

# Quantification of uncertainty from high-dimensional experimental data

Lionel Mathelin <sup>a,b</sup>

<sup>a</sup>*LIMSI-CNRS, BP 133, 91403 Orsay, France.*

<sup>b</sup>*Dpt. of Aeronautics and Astronautics, Massachusetts Institute of Technology,  
77 Massachusetts Av., Cambridge, MA 02139, USA.*

---

## Abstract

This paper discusses a methodology for determining a functional representation of a random process from a collection of pointwise samples. This situation typically arises when experimental data are used to estimate some quantities in a form suitable for their subsequent use in a numerical simulation as initial and/or boundary conditions. The present work specifically focuses onto random quantities lying in a high dimensional stochastic space in the context of limited amount of information. The proposed approach involves a procedure for the selection of an approximation basis and the evaluation of the associated coefficients. The selection of the approximation basis relies on the *a priori* choice of the High-Dimensional Model Representation format combined with a modified Least Angle Regression technique. The resulting basis then provides the structure for the actual approximation basis, possibly using different functions, more parsimonious and nonlinear in its coefficients. To evaluate the coefficients, both an alternate least squares and an alternate weighted total least squares methods are employed. Examples are provided for the approximation of a random variable in a high-dimensional space as well as the estimation of a random field. Stochastic dimensions up to 100 are considered, with an amount of information as low as about 3 samples per dimension, and robustness of the approximation is demonstrated w.r.t. noise in the dataset. The computational cost of the solution method is shown to scale only linearly with the cardinality of the *a priori* basis and exhibits a  $(N_q)^s$ ,  $2 \leq s \leq 3$ , dependence with the number  $N_q$  of samples in the dataset. The provided numerical experiments illustrate the ability of the present approach to derive an accurate approximation from scarce experimental data even in the presence of noise.

*Key words:* Uncertainty Quantification, Least Angle Regression, High-Dimensional Model Reduction, Total Least Squares, Alternate Least Squares, Polynomial Chaos.

---



---

*Email address:* mathelin@limsi.fr (Lionel Mathelin).

## 1 Introduction

With the growing available computational power, and as more efficient numerical methods become available, domains as diverse as engineering, chemistry, psychometrics, medicine, finance or social sciences, now heavily rely on simulation for the prediction of more and more complex phenomena, often combining multi-models and high accuracy requirement. The prediction capability of modern simulations is often such that a new bottle-neck for accuracy has emerged from the lack of relevant boundary and/or initial conditions (BICs). Indeed, boundary and initial conditions are often poorly known and have to be estimated or modeled. This introduces modeling errors which often constitute the main source of lack of accuracy in the simulation chain. This situation has triggered a renewed interest for stochastic modeling where it is explicitly accounted for uncertainty in the model. With probabilistic BICs, the output of the simulation is a probabilistic quantity which interpretation and comparison with real data is left to the final user. The BICs may sometimes be modeled from first principles but are often approximated in a functional form involving a set of influencing parameters and identified from experimental measurements. However, more often than not, only relatively few measurements are available, in particular when a significant number of parameters is of influence so that representing the BICs takes the form of a high-dimensional approximation problem.

If the random process, which output is to be represented in closed-form, is driven by known equations, efficient techniques may be used to determine its representation. In the specific case of high-dimensional quantities, tensor-based representations have proved to be effective when applicable. In particular, low-rank approximations based on an *a priori* chosen separated representation can be efficiently derived, see Nouy (2007, 2010*a,b*); Matthies & Zander (2012) in the context of uncertainty quantification (UQ). If a closed-form model description of the process at hand is not available, one is typically left with approximating it from a finite collection of instances, hereafter termed samples. When the process is known only from a closed numerical code used as a black-box or if measurements can be made arbitrarily (design of experiments), some properties of approximation theory can be exploited. For instance, measurements may be taken at some particular locations in the parameter space, possibly associating a weight to them, so that the random Quantity of Interest (QoI) can be represented in the retained approximation basis with good accuracy using (sparse) quadrature techniques, Novak & Ritter (1999); see also Xiu & Hesthaven (2005) for an application to UQ. Anisotropy in the QoI may be exploited by biasing the quadrature weights, Nobile *et al.* (2007); Ganapathysubramanian & Zabaras (2007); Ma & Zabaras (2010). In Doostan & Iaccarino (2009), an Alternate Least Squares (ALS) technique to estimate the coefficients has been considered with samples lying on a tensor-product grid. Another situation of design of experiment arises in importance sampling where the Markov-Chain Monte Carlo algorithm requires a new sample at a specific proposed location. This control over the samples usually brings efficiency and allows to approximate a reasonably behaved QoI with accuracy.

A different situation occurs when the data are provided ‘as is’, with no ability to choose

the set of samples nor to add a measurement. This is a common situation, typically arising when measurements come from a past experiment or are costly to acquire so that new samples cannot be taken. In this context, one has to resort to a regression-based approach and the coefficients of the approximation are then solution of an optimization problem. This type of approach was considered in Choi *et al.* (2004); Berveiller *et al.* (2006).

In the present work, the focus is specifically put on deriving a closed-form approximation of a high-dimensional quantity of interest from a small, uncontrolled, collection of its samples. No governing equation is available and the dataset of samples is the only piece of information one can rely on. This requires to determine an approximation basis finely tuned to the data at hand and an efficient way of evaluating the associated coefficients. To this aim, we rely on the fact that, as a counterpart of the curse of dimensionality associated with high-dimensional problems, real applications often reward with a *blessing* of dimensionality. Indeed, in many cases, the QoI can be well approximated in a low-dimensional subspace of the solution space, sometimes involving orders of magnitude fewer degrees-of-freedom. This typically occurs when the solution exhibits some degree of sparsity in the retained functional space. Efficient techniques have been proposed in the recent past to take advantage of this and essentially consist in matching the approximation with the observational data while promoting a sparse coefficient set. This class of methods work well in many different contexts and have been recently applied to the UQ framework, Doostan & Owhadi (2011); Mathelin & Gallivan (2012). These techniques rely on the Compressed Sensing theory, *e.g.*, Candès & Tao (2004a); Donoho (2006), and may seem well suited for the present problem as they promote a low cardinality approximation of the QoI. However, they require to handle a potentially huge representation basis, or *dictionary*, and associated optimization problem, leading to severe memory and computation limitations in the present high-dimensional context.

In this paper, we explore a solution method combining the strength of different techniques, taking advantage of the sparsity of the representation in a suitable basis and allowing an efficient approximation of a well-behaved multivariate function with a low number of degrees-of-freedom hence compatible with a small experimental dataset. The driving principle is first to consider a tight approximation basis based on *a priori* knowledge on the QoI at hand and to rely on the available data to further refine it. In a nutshell, an initial approximation basis is first considered in the High-Dimensional Model Representation format (HDMR, Rabitz & Alış (1999); Alış & Rabitz (2001)), assuming it is suitable for representing the QoI. This initial basis is referred to as *a priori*, or *proposal*, basis. Next, available data are used to refine it by retaining only its most relevant basis functions through a constructive subset selection procedure based on a modification of the Least Angle Regression approach proposed in Efron *et al.* (2004). This determines a skeleton, or final approximation basis, and the associated coefficients are then evaluated with an alternate least squares technique. The solution method allows to approximate random variables as well as random fields.

The paper is organized as follows. In section 2, we briefly recall standard techniques for deriving a closed-form approximation of a random variable from a finite set of samples.

Section 3 is devoted to a short discussion of different representation formats of functions in high-dimensional spaces. The proposed solution method is introduced and discussed in section 4 along with a simple motivating example. An algorithm is proposed and numerical results are provided, demonstrating its effectiveness. Scalability of the proposed approach together with its robustness w.r.t. noise in the data is also discussed. In section 5, the present methodology is illustrated on a stochastic diffusion equation involving up to 100 dimensions and on the space-dependent solution of the Shallow Water Equations with random parameters. Accuracy, robustness and scalability of the proposed approach are shown. Concluding remarks close the paper in section 6.

## 2 Quantification of uncertainty

We briefly discuss the representation of a random quantity and standard ways of evaluating it in closed-form from a discrete set of samples.

### 2.1 General framework

Random quantities are defined on a probability space  $(\Theta, \mathcal{B}_\Theta, \mu_\Theta)$  where  $\Theta$  is the space of elementary events  $\theta \in \Theta$ ,  $\mathcal{B}_\Theta$  a  $\sigma$ -algebra defined on  $\Theta$  and  $\mu_\Theta$  a probability measure on  $\mathcal{B}_\Theta$ . To make the description of the problem amenable to a tractable representation, it is convenient to introduce a finite set of statistically independent random variables  $\{\xi_i\}_{i=1}^d : \Theta \rightarrow \Xi_i \subseteq \mathbb{R}$ ,  $\theta \mapsto \xi_i(\theta)$ . The set of these  $d$  random variables is defined on a probability space  $(\Xi, \mathcal{B}_\Xi, \mu_\Xi)$  with  $\Xi = \times_{i=1}^d \Xi_i = \boldsymbol{\xi}(\Theta) \subseteq \mathbb{R}^d$ ,  $\boldsymbol{\xi} := (\xi_1 \dots \xi_d)$ ,  $\mathcal{B}_\Xi \subset 2^\Xi$  a  $\sigma$ -algebra on  $\Xi$  and  $\mu_\Xi = \mu_\Theta \circ \boldsymbol{\xi}^{-1}$  the probability measure on  $\mathcal{B}_\Xi$ . Since the physical process at hand relies on random quantities belonging to  $(\Theta, \mathcal{B}_\Theta, \mu_\Theta)$ , a suitable description of its output, or its solution in case the physical process is described by a known mathematical model, may be determined in  $(\Xi, \mathcal{B}_\Xi, \mu_\Xi)$  as justified by the Doob-Dynkin lemma.

In this work, we restrict ourselves to random variables of physical significance, *i.e.*, real-valued second order variables satisfying:

$$\mathcal{E}_\theta [u(\theta)^2] := \int_\Theta u(\theta)^2 d\mu_\Theta(\theta) = \int_\Xi u(\boldsymbol{\xi})^2 d\mu_\Xi(\boldsymbol{\xi}) =: \mathcal{E}_\xi [u(\boldsymbol{\xi})^2] < +\infty, \quad (1)$$

where  $\mathcal{E}$  denotes the expectation operator and  $u$  is the quantity of interest (QoI). It is then natural to consider the space of square integrable functions  $\mathcal{S}$  for describing real-valued functions of the random quantities:

$$\mathcal{S} := L^2(\Xi, \mu_\Xi) = \left\{ v : \Xi \rightarrow \mathbb{R}, \boldsymbol{\xi} \mapsto v(\boldsymbol{\xi}); \mathcal{E}_\xi [v(\boldsymbol{\xi})^2] < +\infty \right\}. \quad (2)$$

Upon introduction of a natural inner product of  $\mathcal{S}$ :  $\langle v, w \rangle_{L^2(\Xi, \mu_\Xi)} := \int_\Xi v(\boldsymbol{\xi}) w(\boldsymbol{\xi}) d\mu_\Xi(\boldsymbol{\xi})$ ,

$\forall v, w \in \mathcal{S}$ , and the associated norm  $\|v\|_{L^2(\Xi, \mu_\Xi)}^2 := \langle v, v \rangle_{L^2(\Xi, \mu_\Xi)}$ ,  $\mathcal{S}$  is a Hilbert space. Further, we define  $\langle v \rangle_{L^2(\Xi, \mu_\Xi)} := \mathcal{E}_\xi[v(\xi)]$ . One can now rely on functional analysis results and take advantage of approximation theory techniques to characterize the output  $u$ . Introducing a Hilbertian basis  $\{\psi_k\}_{k \in \mathbb{N}}$  of  $\mathcal{S}$ , the output can then be uniquely represented as

$$u(\xi) = \sum_{\alpha} c_{\alpha} \psi_{\alpha}(\xi). \quad (3)$$

The basis  $\{\psi_{\alpha}\}_{\alpha \in \mathbb{N}}$  is typically chosen orthonormal w.r.t. the inner product  $\langle v, w \rangle_{L^2(\Xi, \mu_\Xi)}$ . The orthonormality of the basis leads to  $\langle \psi_{\alpha}, \psi_{\alpha'} \rangle_{L^2(\Xi, \mu_\Xi)} = \delta_{\alpha\alpha'}$ ,  $\forall \alpha, \alpha' \in \mathbb{N}$ , with  $\delta$  the Kronecker delta, and the decomposition coefficients  $\{c_{\alpha}\}$  then express as

$$c_{\alpha} = \langle u, \psi_{\alpha} \rangle_{L^2(\Xi, \mu_\Xi)} = \int_{\Xi} u(\xi) \psi_{\alpha}(\xi) d\mu_{\Xi}(\xi), \quad \forall \alpha \in \mathbb{N}. \quad (4)$$

For a given representation basis  $\{\psi_{\alpha}\}$  of  $\mathcal{S}$ , the output  $u(\xi)$  is entirely characterized by the set of coefficients  $\{c_{\alpha}\}$ . For computational purpose, the infinite dimensional representation Eq. (3) is substituted with a finite dimensional approximation relying on a subset  $\mathcal{J} \subset \mathbb{N}$  of the representation basis:

$$u(\xi) \approx \sum_{\alpha \in \mathcal{J}} c_{\alpha} \psi_{\alpha}(\xi). \quad (5)$$

## 2.2 Computing a data-driven approximation

As seen above, in many situations, a closed-form model of the QoI is not available or not reliable enough to be used and one can only rely on the sole available input-output information to approximate the output  $u$ . The solution method then consists in using a set of outputs given some inputs, *i.e.*, samples of the process. One then looks for a functional form of the map between the set of random variables  $\xi^{(q)}$  and the output value  $u(\xi^{(q)}) =: u^{(q)}$ ,  $\forall 1 \leq q \leq N_q$ , where  $N_q$  is the size of the available experimental set. Approximating the output under the functional form of Eq. (5) results in evaluating the coefficients  $\{c_{\alpha}\}$  from  $\left\{(\xi^{(q)}, u^{(q)})\right\}_{q=1}^{N_q}$ ,  $\xi^{(q)} = (\xi_1^{(q)} \dots \xi_d^{(q)})$ .

### 2.2.1 Monte Carlo

If the sampling can be controlled, in the sense that samples can be drawn arbitrarily, the popular Monte Carlo approach can be followed and the approximation coefficients are then estimated from

$$c_{\alpha} = \int_{\Xi} u(\xi) \psi_{\alpha}(\xi) d\mu_{\Xi}(\xi) \approx \sum_q u(\xi^{(q)}) \psi_{\alpha}(\xi^{(q)}). \quad (6)$$

Monte Carlo-based estimation is very robust and easy to implement but suffers from a slow  $\mathcal{O}(N_q^{-1/2})$  asymptotic convergence rate. However, since the convergence rate does not depend on the dimensionality of the integral, this is a wise choice for very high-dimensional problems where other methods fail. Alternatively, quasi-Monte Carlo methods generate a low-discrepancy sequence of samples improving the convergence rate of the evaluation for moderate- to high-dimensional problems.

### 2.2.2 Quadrature

For low to moderate dimensionality problems, the  $d$ -dimensional integral arising in Eq. (4) may be advantageously evaluated with a quadrature rule:

$$c_\alpha = \int_{\Xi} u(\boldsymbol{\zeta}) \psi_\alpha(\boldsymbol{\zeta}) d\mu_{\Xi}(\boldsymbol{\zeta}) \approx \sum_q w^{(q)} u(\boldsymbol{\xi}^{(q)}) \psi_\alpha(\boldsymbol{\xi}^{(q)}), \quad (7)$$

where  $\{w^{(q)}\}$  are the weights associated with the quadrature points  $\{\boldsymbol{\xi}^{(q)}\}$ . For instance, the Gauss quadrature rule uses only  $N_q$  samples to exactly integrate a polynomial integrand of degree less or equal to  $2N_q - 1$  while the Clenshaw-Curtis is nested, allowing to sequentially increase the order of the quadrature rule without having to re-evaluate points sampled at the previous step.

### 2.2.3 Regression

The above methods require some kind of control over the samples. However, experimental data are often obtained with little, if any, control on the samples, hence no experimental design may be exploited. A solution method is then to reformulate the evaluation of the coefficients as a minimization problem:

$$\mathbf{c} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}|}} \|\mathbf{u} - \Psi \tilde{\mathbf{c}}\|_2, \quad (8)$$

with  $\mathbf{c} = (c_1 \dots c_{|\mathcal{J}|})^T$ ,  $\mathbf{u} = (u^{(1)} \dots u^{(N_q)})^T$ ,  $\Psi \in \mathbb{R}^{N_q \times |\mathcal{J}|}$ ,  $\Psi_{q\alpha} = \psi_\alpha(\boldsymbol{\xi}^{(q)})$  and  $|\mathcal{J}|$  the cardinality of the approximation basis  $\{\psi_\alpha\}_{\alpha \in \mathcal{J}}$ . For a full column rank  $\Psi$ , the solution is given by  $\mathbf{c} = \Psi^+ \mathbf{u}$  which is typically evaluated using the Cholesky decomposition of the symmetric positive definite matrix  $\Psi^T \Psi$  or the QR decomposition of  $\Psi$ . When the size of the dataset grows, this standard Least Squares (LS) problem may become computationally involved. The quasi-regression solution alleviates the computational burden and is given by

$$c_\alpha = \boldsymbol{\psi}_\alpha^T \mathbf{u} / \|\boldsymbol{\psi}_\alpha\|_2^2, \quad \boldsymbol{\psi}_\alpha = \left( \psi_\alpha(\boldsymbol{\xi}^{(1)}) \dots \psi_\alpha(\boldsymbol{\xi}^{(N_q)}) \right)^T, \quad 1 \leq \alpha \leq |\mathcal{J}|. \quad (9)$$

Standard least squares formulation as considered in Eq. (8) treats all predictors  $\{\psi_\alpha\}_{\alpha=1}^{|\mathcal{J}|}$  the same way and uses the available data to estimate all the coefficients to produce

an estimate with a low bias but often a large variance. As will be discussed in section 4.4.1, additional properties of the QoI may be exploited or imposed to the approximation coefficients. This class of approaches trades some increase in bias with a decrease in variance and often results in an improved accuracy. A suitable solution method then typically formulates as a penalized least squares problem:

$$\mathbf{c} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}|}} \|\mathbf{u} - \Psi \tilde{\mathbf{c}}\|_2 + \mathcal{J}(\tilde{\mathbf{c}}). \quad (10)$$

The properties of the penalized LS solution are driven by the choice of the function  $\mathcal{J}$  which flexibility leads to a variety of solution techniques, see Hesterberg *et al.* (2008); Hastie *et al.* (2009). Since we have no control over the sampling strategy, we will rely on regression to estimate the approximation coefficients. The discussion of an efficient least squares formulation in the present context is postponed to section 4.

### 3 Functional representation of random variables

#### 3.1 Polynomial Chaos

As seen above, a random quantity is conveniently approximated in a Hilbertian basis  $\{\psi_k\}$ . This allows for a closed-form representation and is suitable for subsequent use, say, as an input of a numerical model. Many choices may be made for the basis and this brings flexibility. If the random quantity is known, or expected, to exhibit a certain degree of smoothness along the stochastic space, a suitable and popular choice is to take advantage of this smoothness using a spectral-based approximation relying on polynomials. Early efforts towards this direction are the pioneering works of Wiener (1938) who used univariate Hermite polynomials  $\psi_\alpha(\xi_i)$  of zero-centered, unit variance, normal random variables  $\xi_i \sim \mathcal{N}(0, 1)$ . These polynomials define an orthogonal basis of  $L^2(\Xi_i, \mu_{\Xi_i})$ ,  $\mu_{\Xi_i} \propto e^{-\frac{1}{2}\xi_i^2}$ . Tensorization of univariate Hermite polynomials  $\psi$  then leads to an orthogonal basis of  $L^2(\Xi, \mu_\Xi)$ :

$$\langle \psi_\alpha, \psi_{\alpha'} \rangle_{L^2(\Xi, \mu_\Xi)} \propto \int_{\Xi} \psi_\alpha(\boldsymbol{\zeta}) \psi_{\alpha'}(\boldsymbol{\zeta}) e^{-\frac{1}{2}(\boldsymbol{\zeta}^T \boldsymbol{\zeta})} d\boldsymbol{\zeta} \propto \delta_{\alpha\alpha'}. \quad (11)$$

This can be extended to polynomials orthogonal with respect to different measures, Ghanem & Spanos (2003); Xiu & Karniadakis (2002); Soize & Ghanem (2004), and constitutes the so-called (generalized) Polynomial Chaos basis. A common practice is to consider an approximation space  $\mathcal{S}_p$  spanned by polynomials of given maximum total degree  $p$ :

$$\mathcal{S}_p = \text{span} \left( \{ \psi_\alpha(\boldsymbol{\xi}) = \psi_{\alpha_1}(\xi_1) \dots \psi_{\alpha_d}(\xi_d) \}; \boldsymbol{\alpha} = (\alpha_1 \dots \alpha_d), \sum_{i=1}^d \alpha_i \leq p \right), \quad (12)$$

and the number of terms to be determined in the approximation (5) is then  $|\mathcal{J}| =$



$\binom{d+p}{d}$ . We adopt the convention  $\psi_1 \equiv 1$ . When the random quantity is not smooth enough for a low degree polynomial fit to be accurate, approximation schemes such as  $h/p$ -type refinement or Multi-Resolution Analysis may be applied, see Le Maître & Knio (2010).

### 3.2 Tensor-based representation

Some alternative representation formats specifically exploit the tensor-product structure of the Hilbert stochastic space  $\mathcal{S}$ . This class of methods typically approximates a  $d$ -variate function with a series of products of lower dimensional functions. Efficient algorithms allow to determine the approximation coefficients of the representation by solving a series of low-dimensional problems while never considering the full-dimensional problem at once. A general presentation of tensor-structured numerical methods can be found in Khoromskij (2012) while application to the approximation of a high-dimensional random quantity is considered in Khoromskij & Schwab (2011); Nouy (2010b); Matthies & Zander (2012); Doostan & Iaccarino (2009). For instance, a  $d$ -variate quantity may be approximated under a CANDECOMP-PARAFAC (CP) format with a sum of rank-1 terms, the simplest form of tensored-structure format:

$$u(\boldsymbol{\xi}) \approx \sum_{r=1}^{n_r} f_{1,r}(\xi_1) \cdots f_{d,r}(\xi_d), \quad (13)$$

with  $n_r$  the retained rank of the decomposition and  $\{f_{i,r}\}_{i=1}^d$  univariate functions. These univariate functions are usually sequentially determined in turn so that only 1-D problems have to be solved. Assuming  $p$ -th order polynomials for  $\{f_{i,r}\}$ , the resulting cardinality of the approximation is  $d n_r p$ . It thus exhibits a linear dependence with the number of dimensions, in contrast with the exponential dependence of the Polynomial Chaos. Approximation techniques more efficient than the decomposition (13) rely on numerically robust and/or memory efficient data-structures, *e.g.*, Tucker format or Tensor-Trains, see Khoromskij (2012). A tensored-structure format then constitutes a method of choice for deriving memory- and CPU-efficient approximation of high-dimensional quantities. They also lead to a low-cardinality basis  $|\mathcal{J}|$  so that the conditioning of the approximation method remains good, in the sense that  $|\mathcal{J}| \leq N_q$ , a crucial feature for deriving a good approximation from the scarce available data.

### 3.3 High-Dimensional Model Representation

In a similar perspective of using an approximation basis with the least cardinality possible to make the most out of the scarce data, an efficient alternative to the tensored-structure formats of the previous paragraph for representing high-dimensional quantities is discussed



in Rabitz & Alis (1999); Alis & Rabitz (2001). It consists in representing a quantity  $u(\boldsymbol{\xi})$  with a sum of lower-dimensional terms accounting for increasing levels of interaction between the constitutive variables:

$$u(\boldsymbol{\xi}) = f_\emptyset + \sum_{i=1}^d f_i(\xi_i) + \sum_{\substack{i,j=1, \\ j>i}}^d f_{ij}(\xi_i, \xi_j) + \dots + f_{12\dots d}(\xi_1, \dots, \xi_d) = \sum_{\boldsymbol{\gamma} \subseteq \{1, \dots, d\}} f_{\boldsymbol{\gamma}}, \quad (14)$$

where  $f_{\boldsymbol{\gamma}}$  are functions of  $\mathcal{S}$  and depend only on a subset of variables  $\boldsymbol{\xi}_{\boldsymbol{\gamma}} = \{\xi_i\}_{i \in \boldsymbol{\gamma}}$  and  $\boldsymbol{\gamma}$  is a multi-index. This decomposition is exact and does not introduce any approximation. It is also unique. An important property is that the modes  $\{f_{\boldsymbol{\gamma}}\}$  are mutually orthogonal:  $\langle f_{\boldsymbol{\gamma}}, f_{\boldsymbol{\gamma}'} \rangle_{L^2(\Xi, \mu_{\Xi})} = 0$ ,  $\forall \boldsymbol{\gamma} \neq \boldsymbol{\gamma}' \subseteq \{1, \dots, d\}$ . The zero-th order term  $f_\emptyset$  accounts for the mean and is invariant across the entire domain  $\Xi$ , while the other modes are zero-mean:

$$f_\emptyset = \langle u \rangle_{L^2(\Xi, \mu_{\Xi})}, \quad \langle f_{\boldsymbol{\gamma}} \rangle_{L^2(\Xi, \mu_{\Xi})} = 0, \quad \forall \boldsymbol{\gamma} \subseteq \{1, \dots, d\} \setminus \emptyset. \quad (15)$$

First order terms  $\{f_i\}_{i=1}^d$  are responsible for the approximation of  $u(\boldsymbol{\xi})$  with univariate functions of variables acting independently, second-order terms  $\{f_{ij}\}_{i,j=1, j>i}^d$  represent the joint-impact of two variables  $\xi_i$  and  $\xi_j$  onto the output  $u$ , etc. The rationale behind the expected success of this so-called *High Dimensional Model Representation* (HDMR) is that many quantities of interest exhibit a significant dependence on low-dimensional groups of variables only, hence having negligible high order interaction decomposition terms. This leads to an efficient approximation of  $u$  with only a low  $N_l$ -order HDMR:  $u(\boldsymbol{\xi}) \approx \sum_{\boldsymbol{\gamma} \subseteq \{1, \dots, d\}} f_{\boldsymbol{\gamma}}(\boldsymbol{\xi}_{\boldsymbol{\gamma}})$ ,  $|\boldsymbol{\gamma}| \leq N_l$ . We denote  $\mathcal{J}_f$  the set of retained modes,  $\mathcal{J}_f := \{\boldsymbol{\gamma} \subseteq \{1, \dots, d\}; |\boldsymbol{\gamma}| \leq N_l\}$ .

Functions  $\{f_{\boldsymbol{\gamma}}\}$  are evaluated with the application of a set of commuting projections  $\{\mathcal{P}_i\}$  onto the output  $u$ . The projection  $\mathcal{P}_i$  eliminates the effect of variable  $\xi_i$  while leaving the effect of the others unchanged. Letting  $\mathcal{P}_\emptyset$  be the identity operator on  $\mathcal{S}$ , we define  $\mathcal{P}_\boldsymbol{\eta} = \prod_{i \in \boldsymbol{\eta}} \mathcal{P}_i$ ,  $\forall \boldsymbol{\eta} \subseteq \{1, \dots, d\}$ . Functions  $\{f_{\boldsymbol{\gamma}}\}$  can then be written, Kuo *et al.* (2009),

$$f_{\boldsymbol{\gamma} \subseteq \{1, \dots, d\} \setminus \emptyset} = \mathcal{P}_{\{1, \dots, d\} \setminus \boldsymbol{\gamma}} u - \sum_{\boldsymbol{\gamma}' \subsetneq \boldsymbol{\gamma}} f_{\boldsymbol{\gamma}'} = \sum_{\boldsymbol{\gamma}' \subsetneq \boldsymbol{\gamma}} (-1)^{|\boldsymbol{\gamma}| - |\boldsymbol{\gamma}'|} \mathcal{P}_{\{1, \dots, d\} \setminus \boldsymbol{\gamma}'} u, \quad f_\emptyset = \mathcal{P}_{\{1, \dots, d\}} u. \quad (16)$$

Defining projections as  $\mathcal{P}_i u(\boldsymbol{\xi}) = \int_{\Xi_i} u(\xi_1, \dots, \xi_{i-1}, \zeta', \xi_{i+1}, \dots, \xi_d) d\mu(\zeta')$ , the measure  $\mu$  determines the form of the projection. A popular choice consists in using  $\mu = \mu_{\Xi_i}$  so that the *Analysis of Variance* (ANOVA) decomposition is obtained. An example of application of the HDMR representation to the approximation of a random quantity is presented in Ma & Zabarar (2010).

## 4 Quantifying the uncertainty of experimental data

### 4.1 Setting up the stage

In the following, we will consider that the quantity of interest  $u$  is a scalar-valued random field, indexed by space and/or time  $\mathbf{x} \in \mathbb{R}^{d_x}$  and depending on a set of random variables  $\boldsymbol{\xi} \in \mathbb{R}^d$ . To approximate it, the only available piece of information is the collection of samples  $\{\mathbf{x}^{(q)}, \boldsymbol{\xi}^{(q)}, u^{(q)}\}_{q=1}^{N_q}$ . Since the underlying random quantity is only known through these samples, no governing equation for the QoI can be exploited and, say, Galerkin projection-based weak-formulation methods cannot be employed. Further, these samples are collected at random and do not follow a deterministic rule so that no deterministic sampling strategy can be assumed. Quadrature-based techniques can then not be applied either and one has to resort to regression to estimate the coefficients of the approximation in the retained basis  $\{\psi_\alpha\}$ . Standard  $L^2$ -regression solves Eq. (8) which is only well-posed for a matrix  $\Psi$  such that  $\Psi^T \Psi$  is invertible so that it requires the number of observations to be larger than the cardinality of the approximation basis,  $N_q \geq |\mathcal{J}|$ . Since the samples are believed to lie in a high-dimensional space selecting an approximation basis which both leads to a well-posed regression problem and a good representation accuracy could be a daunting task when observational data is limited.

The choice of a good approximation basis in a general setting largely remains an open question. On one hand, if one is given a dictionary of approximation functions, *a priori* selecting the best terms so that they can be evaluated from the data is generally a NP-hard problem. On the other hand, dictionary-learning techniques require a training while availability of an independent training set cannot be assumed here.

We first consider the approximation strategy for a random process in the form of a separated representation  $u(\mathbf{x}, \boldsymbol{\xi}) \approx \sum_n w_n(\mathbf{x}) \lambda_n(\boldsymbol{\xi})$ , section 4.2. Modes  $\{w_n(\mathbf{x})\}$  are associated with all physical dimensions the random process may be indexed upon (space, time, ...) so that  $\mathbf{x} = (x_1 x_2 \dots t \dots) \subseteq \mathbb{R}^{d_x}$ . Examples considered in this paper are space-dependent random variables so that  $\mathbf{x} = (x_1 x_2 \dots x_{d_x})$  and  $\{w_n(\mathbf{x})\}$  will be referred to as ‘spatial modes’. In contrast to  $\mathbf{x}$ , the dimension of  $\boldsymbol{\xi}$  can be very large and a specific strategy is then discussed for representing the ‘stochastic’ modes  $\{\lambda_n(\boldsymbol{\xi})\}$ , sections 4.3–4.5.

In the present work, we separate the determination of an efficient representation format from the evaluation of the coefficients. We first choose an *a priori*, or *proposal*, format for the approximation, section 4.3, and leave the selection of particular terms to be included in the retained approximation basis to a dedicated subset selection procedure which will further refine the approximation basis, section 4.4. Results from Compressed Sensing show that the number of samples necessary for accurately selecting the dominant basis functions of a  $K$ -sparse QoI varies as  $K \log(|\mathcal{J}|)$ , Candès & Romberg (2006), illustrating the fact that it becomes increasingly difficult to select the best terms when the size  $|\mathcal{J}|$  of the proposal dictionary increases. Hence, for the subset selection step to be efficient in terms

of required amount of data, the retained *a priori* format has to be tight, in the sense of leading to a low approximation error with a low number of degrees of freedom (terms): it should be chosen as tight as possible based on knowledge one may have on the QoI at hand. As will be seen below, the computational cost associated with the subset selection scales with the cardinality of the proposal dictionary. The subset selection hence produces a basis, or skeleton, suitable for the data at hand. However, it requires a basis *linear* in its predictors. For this reason, once a set of functions found contributing significantly to the approximation is determined, the actual approximation may be evaluated with a different basis, of the same skeleton, possibly nonlinear in its predictors such as a tensor-based approximation. Evaluation of the final approximation is discussed in section 4.5.

## 4.2 Approximation of a random process

The approximation of a random process, say, a space-dependent uncertain quantity  $u(\mathbf{x}, \boldsymbol{\xi})$  is considered. An approximation is determined in the form of separation of variables:

$$u(\mathbf{x}, \boldsymbol{\xi}) \approx w_0(\mathbf{x}) + \sum_{n=1}^N w_n(\mathbf{x}) \lambda_n(\boldsymbol{\xi}). \quad (17)$$

For convenience, the decomposition is rewritten as  $u(\mathbf{x}, \boldsymbol{\xi}) \approx \sum_{n=0}^N w_n(\mathbf{x}) \lambda_n(\boldsymbol{\xi})$ , with  $\lambda_0 \equiv 1$ . The spatial modes are defined as:  $w_n(\mathbf{x}) = \sum_{l=1}^{|\mathcal{J}_x|} c_{l,n}^{(w)} \phi_l(\mathbf{x})$  with  $\{\phi_l\}$  a chosen truncated basis of cardinality  $|\mathcal{J}_x|$ . The functional form of stochastic modes  $\{\lambda_n\}$  will be discussed in section 4.3.

The spatial and stochastic modes of the approximation (17) are sequentially determined in turn. Let  $\|v\|_{N_q} := \langle v, v \rangle_{N_q}$  be the norm induced by the data-driven inner product:  $\langle \cdot, \cdot \rangle_{N_q} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $(v, w) \mapsto \langle v, w \rangle_{N_q} := \sum_{q=1}^{N_q} v^{(q)} w^{(q)}$ . Assuming  $\{\lambda_n\}$  known and projecting Eq. (17) onto the space spanned by  $\{\phi_l\}$ , the coefficients  $\{c_{l,n}^{(w)}\}_l$  of the deterministic mode  $w_n$  are the solution of the following problem:

$$\begin{aligned} \langle u, \phi_k \lambda_n \rangle_{N_q} &= \left\langle \sum_{n'=0}^{n-1} w_{n'} \lambda_{n'} + \sum_{l=1}^{|\mathcal{J}_x|} c_{l,n}^{(w)} \phi_l \lambda_n, \phi_k \lambda_n \right\rangle_{N_q}, \quad \forall 1 \leq k \leq |\mathcal{J}_x|, \\ \Leftrightarrow \mathbf{c}_{\cdot,n}^{(w)} &= \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}_x|}} \left\| \mathbf{u} - \sum_{n'=0}^{n-1} \mathbf{w}_{n'} \odot \boldsymbol{\lambda}_{n'} - (\Phi \tilde{\mathbf{c}}) \odot \boldsymbol{\lambda}_n \right\|_2, \end{aligned} \quad (18)$$

where  $\Phi \in \mathbb{R}^{N_q \times |\mathcal{J}_x|}$ ,  $\Phi_{ql} = \phi_l(\mathbf{x}^{(q)})$ ,  $\mathbf{w}_n = (w_n(x^{(1)}) \dots w_n(x^{(N_q)}))^T$ ,  $\boldsymbol{\lambda}_n = (\lambda_n(\boldsymbol{\xi}^{(1)}) \dots \lambda_n(\boldsymbol{\xi}^{(N_q)}))^T$  and  $\odot$  is the Hadamard product. Similarly, the stochastic mode  $\lambda_n$  is evaluated by determining the set of coefficients  $\{c_n^{(\lambda)}\}$  minimizing  $\left\| \mathbf{u} - \sum_{n'=0}^{n-1} \mathbf{w}_{n'} \odot \boldsymbol{\lambda}_{n'} - \mathbf{w}_n \odot \boldsymbol{\lambda}_n \left( \{c_n^{(\lambda)}\} \right) \right\|_2$  using Algorithm 2 to be presented in section 4.5 The spatial mode  $w_n$  is then evaluated from Eq. (18) given all the other information and the whole iteration is repeated until

convergence of the pair  $\{w_n, \lambda_n\}$ . The next pair can then be determined with the same methodology with  $n \leftarrow n + 1$ . The algorithm is summarized in Algorithm 1.

---

**Algorithm 1** Sketch of the solution method for approximating a random process

---

- 1: Choose  $|\mathcal{J}_x|$ . Initialize  $\mathbf{z} = \mathbf{u}$  and  $\boldsymbol{\lambda}_0 = \mathbf{1}$  and set  $n \leftarrow 0$ .
  - 2: **repeat**
  - 3:   **while**  $\|\lambda_n\|_{N_q}$  not converged **do**
  - 4:     **Solve for coefficients**  $\{c_{l,n}^{(w)}\}_{l=1}^{|\mathcal{J}_x|}$  **of the deterministic mode** given  $\boldsymbol{\lambda}_n$  and  $\mathbf{z}$  using Eq. (18) and normalize  $w_n(\mathbf{x}) = \sum_l c_{l,n}^{(w)} \phi_l(\mathbf{x})$ ,  $\|w_n\|_{N_q} = 1$ .
  - 5:     **Solve for the stochastic mode**  $\lambda_n(\boldsymbol{\xi})$  using Algorithm 2 given  $\mathbf{w}_n$  and  $\mathbf{z}$ . If  $n = 0$ ,  $\boldsymbol{\lambda}_0 \leftarrow \mathbf{1}$ .
  - 6:   **end while**
  - 7:   Set  $\mathbf{z} \leftarrow \mathbf{z} - \mathbf{w}_n \odot \boldsymbol{\lambda}_n$ , and  $n \leftarrow n + 1$ .
  - 8: **until** a termination criterion is met (for instance,  $\|\lambda_n\|_{N_q}$  below a threshold or maximum rank  $n > N$  reached).
- 

**Remark 1** *If the separated approximation grows beyond a few modes, it is beneficial to update the coefficients of, say, the spatial modes for improved accuracy: solve for  $\{w_0, \dots, w_n\}$  given  $\{\mathbf{u}, \lambda_0, \dots, \lambda_n\}$ .*

#### 4.3 A basis suitable for scarce data regression of a random variable

Since the solution method discussed below is not specific to as separated format  $u(\mathbf{x}, \boldsymbol{\xi}) \approx \sum_n w_n(\mathbf{x}) \lambda_n(\boldsymbol{\xi})$  but applies to random variables in general, the QoI  $u$  referred to in the subsequent sections is any stochastic mode  $\lambda_n(\boldsymbol{\xi})$ . The discussion then focuses on approximating a high-dimensional random variable  $u(\boldsymbol{\xi})$ .

In this work, we want to take advantage of the low order interactions of constitutive variables for many quantities of practical interest as mentioned in section 3.3. Previous works have shown evidence of this low interaction configuration in various situations, Rabitz & Alış (1999); Alış & Rabitz (2001); Ma & Zabaras (2010), and the QoI is hence chosen to be approximated under the HDMR form, Eq. (14).

To assess the choice of our *a priori* functional form for approximating a random variable, and while choosing a good basis is problem-dependent, let us consider a simple motivating example in the form of a steady-state stochastic diffusion equation on  $\Omega \times \Xi$ ,  $\Omega = [x_-, x_+] \subset \mathbb{R}$ , with deterministic Dirichlet boundary conditions:

$$\nabla_x (\nu(x, \boldsymbol{\xi}') \nabla_x u(x, \boldsymbol{\xi})) = F(x, \boldsymbol{\xi}''), \quad u(x_-, \boldsymbol{\xi}) = u_-, \quad u(x_+, \boldsymbol{\xi}) = u_+. \quad (19)$$

The right-hand side  $F$  is a random source field and  $\nu$  is a space-dependent random diffusion coefficient modeled as:

$$\begin{aligned}\nu(x, \boldsymbol{\xi}') &= \nu_0(x) + \sum_{k=1}^{d_\nu} \sqrt{\sigma_{\nu,k}} \omega_{\nu,k}(x) \xi'_k, & \boldsymbol{\xi}' &= (\xi'_1 \dots \xi'_{d_\nu}), \\ F(x, \boldsymbol{\xi}'') &= F_0(x) + \sum_{k=1}^{d_F} \sqrt{\sigma_{F,k}} \omega_{F,k}(x) \xi''_k, & \boldsymbol{\xi}'' &= (\xi''_1 \dots \xi''_{d_F}),\end{aligned}\tag{20}$$

with  $\nu_0(x) = 1$  and  $F_0(x) = -1$  the respective mean values. The random variables  $\{\xi'_1, \dots, \xi'_{d_\nu}, \xi''_1, \dots, \xi''_{d_F}\}$  are chosen mutually independent and uniformly distributed on  $[0, 1]$ . The spatial modes  $\omega_{\nu,k}(x)$  and  $\omega_{F,k}(x)$ , and their associated amplitude  $\sqrt{\sigma_{\nu,k}}$  and  $\sqrt{\sigma_{F,k}}$ , are the first dominant eigenfunctions of the following eigenproblems:

$$\begin{aligned}\int_{\Omega} K_{\nu}(x, x') \omega_{\nu,k}(x') dx' &= \sigma_{\nu,k} \omega_{\nu,k}(x), & K_{\nu}(x, x') &= \sigma_{\nu}^2 e^{-\frac{(x-x')^2}{2L_{c,\nu}^2}}, \\ \int_{\Omega} K_F(x, x') \omega_{F,k}(x') dx' &= \sigma_{F,k} \omega_{F,k}(x), & K_F(x, x') &= \sigma_F^2 e^{-\frac{(x-x')^2}{2L_{c,F}^2}},\end{aligned}\tag{21}$$

with  $K_{\nu}$  and  $K_F$  the correlation kernels. The random fields properties are chosen as  $\sigma_{\nu} = 0.7$ ,  $\sigma_F = 0.7$ ,  $L_{c,\nu} = 0.3$ ,  $L_{c,F} = 0.3$ . Note that  $\nu(\cdot, \boldsymbol{\xi}') > 0$  a.e., and  $F(\cdot, \boldsymbol{\xi}'') < 0$  a.e., so that the problem remains coercive. The spectra of the operators associated with these eigenproblems are here the same and decay quickly thanks to the high correlation length as can be appreciated from Table 1 where the dominant eigenvalues are given. The resulting problem is then anisotropic in  $\Xi$  in the sense that the degree of dependence of the input random parameters along the dimensions  $\{\xi_1, \dots, \xi_8\}$  strongly varies.

$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$	$\sigma_6$	$\sigma_7$	$\sigma_8$
0.1815	0.1396	0.0906	0.0450	0.0236	0.0097	0.0035	0.0011

Table 1

Upper part of the spectrum of eigenproblems (21).

Denoting  $\boldsymbol{\xi} = (\boldsymbol{\xi}' \boldsymbol{\xi}'') \in \mathbb{R}^d$ ,  $d = d_{\nu} + d_F$ , the solution  $u$  is approximated in a rank- $N$  separated form:  $u(x, \boldsymbol{\xi}) \approx \hat{u}(x, \boldsymbol{\xi}) = w_0(x) + \sum_{n=1}^N w_n(x) \lambda_n(\boldsymbol{\xi})$ . Each stochastic mode  $\lambda_n(\boldsymbol{\xi})$  is determined either in a CP-like format, Eq. (13), or as a HDMR decomposition Eqs. (14). In the latter case, interaction modes  $\{f_{\gamma}\}$  are approximated with a low-rank canonical decomposition on tensorized, unit-normed, univariate polynomials of maximum degree  $p$ :  $f_{\gamma}(\{\xi_i\}_{i \in \gamma}) \approx \sum_{r=1}^{n_r} \prod_{i \in \gamma} \sum_{\alpha=2}^p c_{\gamma,\alpha}^{r,i} \psi_{\alpha}(\xi_i)$ . This approximation is hereafter referred to as a CP-HDMR decomposition. Similarly, univariate functions  $\{f_{i,r}\}_{i=1}^d$  involved in the CP decomposition Eq. (13) are approximated with the same polynomials:  $f_{i,r}(\xi_i) \approx \sum_{\alpha=1}^p c_{\alpha,i,r} \psi_{\alpha}(\xi_i)$ .

The representation of the stochastic modes here relies on  $p = 8$ -th order univariate Legendre polynomials  $\{\psi_{\alpha}\}$ . The dimension of the problem is chosen to be  $d_{\nu} = d_F = 5$  so that  $d = 10$ . The approximation accuracy is estimated by the relative  $L^2$ -norm  $\varepsilon$  of the approximation error estimation evaluated from a  $\widetilde{N}_q$ -point test set  $\left\{ \left( x^{(\tilde{q})}, \boldsymbol{\xi}^{(\tilde{q})}, u^{(\tilde{q})} \right) \right\}_{\tilde{q}=1}^{\widetilde{N}_q}$ ,

independent from the training set:

$$\varepsilon^2 := \frac{\|\mathbf{u} - \hat{\mathbf{u}}\|_2^2}{\|\mathbf{u}\|_2^2}, \quad \mathbf{u} = \left( u^{(1)} \dots u^{(\widetilde{N}_q)} \right), \quad \hat{\mathbf{u}} = \left( \hat{u}^{(1)} \dots \hat{u}^{(\widetilde{N}_q)} \right). \quad (22)$$

The CP-HDMR expansion is here built sequentially, starting with 0-th and 1-st order interaction modes only. From this first approximation of the output, the set of dominant dimensions is estimated from the  $L^2$ -norm of univariate interaction modes  $\{f_\gamma\}_{|\gamma|=1}$ . Only second order interaction modes  $\{f_\gamma\}_{|\gamma|=2}$  in these dominant dimensions are next estimated and the set of dominant dimensions is then further refined based on both 1-st and 2-nd order interaction modes via the sensitivity Sobol indices, see section 4.8. Third order modes are then computed for this new set of dominant dimensions only and the procedure is repeated until some stopping criterion is met, for instance a maximum interaction order  $N_l$  or a maximum basis cardinality  $|\mathcal{J}|$ . The number of samples  $N_q$  is here chosen sufficiently large so that full knowledge on  $u$  can be assumed. The approximation error then only comes from the choice of the approximation basis format, allowing a comparison. This section is loose on details, focusing on the main conclusions and leaving more in-depth discussion for subsequent sections.

First, the accuracy of the CP-HDMR approximation as a function of the decomposition rank  $N$  is studied in terms of  $\varepsilon$ , Eq. (22). Plotted in Fig. 1, the approximation error estimation  $\varepsilon$  decreases when the maximum interaction order  $N_l$  increases from 1 to 3 and as the decomposition rank  $N$  increases.

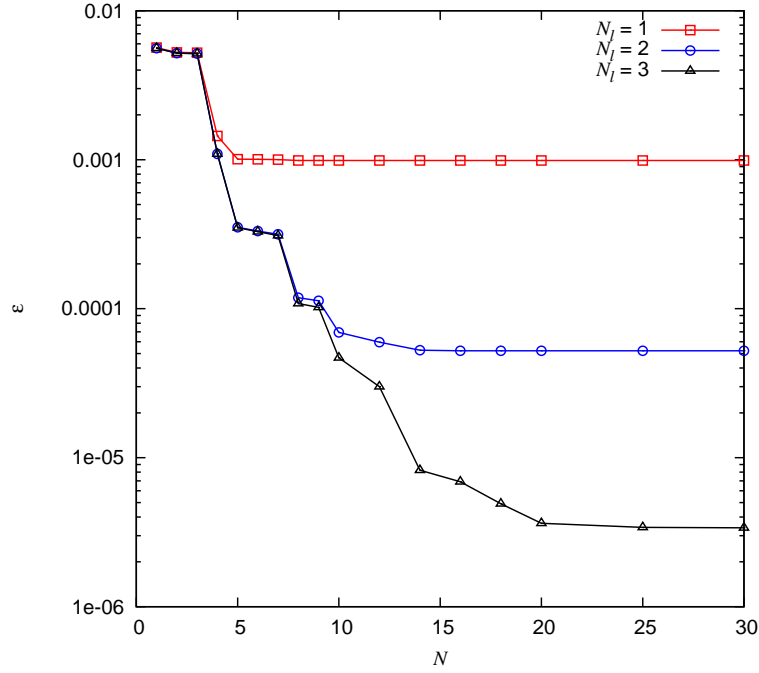


Fig. 1. Convergence of the error estimation of the approximation with the order  $N$  of the separated representation and the maximum interaction order  $N_l$  of the HDMR expansion.

The approximation is seen to improve exponentially fast as the number of modes  $N$  in the

separated representation increases until it reaches a plateau at the best approximation one can get with the retained interaction order  $N_l$ . Increasing the interaction order leads to an improved approximation: increasing from first to second order brings more than a one-order of magnitude improvement in the approximation error estimation and an additional order of magnitude from  $N_l = 2$  to  $N_l = 3$ . In this  $d = 10$  example, the approximation hence exhibits a high convergence rate with  $N_l$ , supporting our assumption that low-order interactions dominate the HDMR decomposition.

This CP-HDMR approximation of the stochastic modes is now compared with a CP-like approximation in the form of Eq. (13), derived with an algorithm similar to that in Nouy (2010b). Both decompositions rely on the same approximation basis for the deterministic modes  $\{w_n\}$ . We focus on the accuracy of the reconstruction as a function of the cardinality of the whole approximation basis both for  $d = 10$  and  $d = 40$  when the maximum decomposition rank  $N$  varies, see Fig. 2. The total cardinality increases as more terms are considered in the decomposition series.

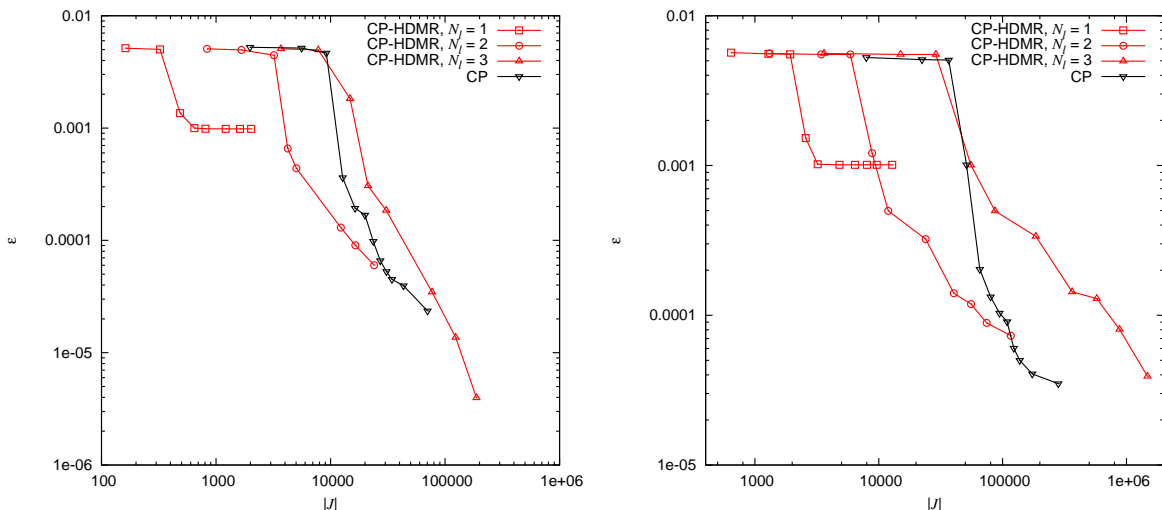


Fig. 2. Convergence of the approximation error estimation  $\varepsilon$  with the total cardinality  $|\mathcal{J}|$  of the representation basis. Approximations of the stochastic modes with the CP-HDMR and CP-like format are compared.  $d = 10$  (left) and  $d = 40$  (right).

The accuracy of the representation is seen to improve as more terms are considered in the series expansion and both the CP and the CP-HDMR formats exhibit an exponential convergence with the total size  $|\mathcal{J}|$  of the decomposition. The CP decomposition achieves a  $10^{-4}$  relative error with a mere 20,000 coefficients ( $d = 10$ ) and with about 100,000 for  $d = 40$ . The first order CP-HDMR quickly reduces the error but plateaus as the functional space is small. A second order approximation already allows a similar  $10^{-4}$  relative error with about 10,000 unknowns for  $d = 10$  and 100,000 for  $d = 40$ . The third order CP-HDMR decomposition is more costly in number of coefficients to determine to reach a given error and, in the present settings, should only be considered if high accuracy is needed.

Unless the targeted accuracy is really high, this motivating example tends to indicate that, for a reasonable required accuracy, a CP-HDMR format involves fewer unknowns than a



CP-like decomposition, both for a low  $d = 10$ - and a moderate  $d = 40$ -dimensional problem. This is an important point since the number of coefficients which can be evaluated with a reasonable accuracy from experimental data is directly related to the size of the available dataset. Moreover, the Sobol indices-based threshold used to incorporate higher interaction order terms was only coarsely hand-tuned leaving significant room for further improved CP-HDMR performance. Finally, a CP-HDMR format allows a great flexibility in representing interaction modes  $\{f_\gamma\}$ . In particular, a more parsimonious representation will be used in the sequel and will achieve a similar accuracy with a lower number of terms. This motivating example hence demonstrates that a general CP-HDMR format approximation may compare favorably with a tensor-based approximation in terms of required number of basis functions for a given reconstruction accuracy, even for reasonably large dimensional problems, and motivates our choice of an HDMR format for the *a priori* basis.

#### 4.4 Subset selection

We now build up from the *a priori* basis and further improve it with an *a posteriori*, data-driven, procedure.

##### 4.4.1 A direct approach

As discussed in section 2.2, different techniques may be used to compute the coefficients of an approximation. In the case considered in this paper, they are evaluated with a data-driven approach from the  $N_q$  pieces of information available on the QoI. However, the available data is scarce while the cardinality  $|\mathcal{J}_{\text{prior}}|$  of the proposal approximation basis may be large, in particular when the dimensionality  $d$  of the problem is large. It can then result in an ill-posed problem where one has to estimate  $|\mathcal{J}_{\text{prior}}|$  coefficients for each stochastic mode  $\lambda_n$  from  $N_q \ll |\mathcal{J}_{\text{prior}}|$  pieces of information. However, this situation often only reflects our lack of knowledge on the quantity at hand and how conservative this naive approximation method is. Indeed, the “blessing of the dimensionality” yields that high-dimensional problems are often intrinsically sparse and lower dimensional. In the present setting, it is likely that many dimensions actually hardly contribute to the approximation and that representing the dependence of the QoI along only a subset of the dimensions yields an acceptable accuracy. In our HDMR representation, it means that many interaction modes  $\{f_\gamma\}$  can be discarded without significantly affecting the accuracy. The challenge for an efficient solution method is then to reveal and exploit the low-dimensional manifold onto which a good approximation of the solution lies. This leads to a sparse representation where most of the coefficients are negligible w.r.t. the dominant ones, hence yielding a well-posed problem,  $|\mathcal{J}_{\text{prior}}| \leq N_q$ . As an illustration, if  $u(\boldsymbol{\xi}) = g(\xi_i)$  was depending only on one dimension  $i$ ,  $i \in \{1, \dots, d\}$ , information theory allows to show that one only requires  $m + 1 + \lceil \log_2 d \rceil$  function evaluations to approximate a sufficiently smooth function  $g \in C^s$ , having  $s$  continuous derivatives, so that  $\|u - \hat{u}\|_{C(\Xi_i)} \leq a h^s$ ,

$h := 1/m$ , where  $a \geq 0$  is related to a norm of  $g$ , DeVore *et al.* (2011). This number of samples actually is directly related to the number of information bits required to represent the integer  $i \in [1, d]$ .

While determining which coefficients are dominant is an NP-hard problem in general, recent results have shown that a good estimation of the best subset can be obtained as the solution of a convex optimization problem. In particular, the LASSO formulation has been extensively studied and proved effective, Tibshirani (1996). One formulation of the LASSO, referred to as *Basis Pursuit Denoising*, writes:

$$\mathbf{c} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}_{\text{prior}}|}} \|\tilde{\mathbf{c}}\|_1 \quad \text{s.t.} \quad \|\mathbf{u} - \Psi \tilde{\mathbf{c}}\|_2 \leq \epsilon, \quad (23)$$

with  $\Psi$  the matrix of evaluations of the approximation basis and  $\epsilon$  the approximation residual. Efforts from the signal processing community, where the theory supporting these results is termed *Compressed Sensing*, have demonstrated its good recovery properties in the case where  $N_q < |\mathcal{J}_{\text{prior}}|$ , *e.g.*, Chen *et al.* (1999); Candès & Tao (2004b); Donoho (2006). In particular, this formulation achieves provable and robust recovery bounds: for a sufficiently incoherent set of approximation and test functions, a  $K$ -sparse solution  $\mathbf{c}$  to Eq. (23) satisfies, Cai *et al.* (2010),

$$\|\mathbf{c}^* - \mathbf{c}\|_2 \lesssim h \left( \epsilon + \frac{\|\mathbf{c}^* - \mathbf{c}_K^*\|_1}{\sqrt{K}} \right), \quad (24)$$

where  $h > 0$  is a constant depending on the set of approximation and test functions and  $\mathbf{c}_K^*$  is the  $K$ -term approximation of  $\mathbf{c}^*$  given by an oracle, *i.e.*, it is the best  $K$ -term approximation of  $\mathbf{c}^*$  if one was given full knowledge of it.

The Compressed Sensing technique was proved very effective and is now being applied in many areas, including Uncertainty Quantification, Doostan & Owhadi (2011); Mathelin & Gallivan (2012). However, standard implementations of the algorithm require the sensing matrix  $\Psi$  to be available. This bears an intrinsic limitation when it comes to high-dimensional problems as it requires the use of the whole dictionary at once from which to select the basis functions associated with the dominant coefficients. This can then be seen as a top-to-bottom approach where one starts with a whole set of basis functions and shrinks it so that the best subset remains, in the sense of Eq. (23). While effective, this approach is not tractable for high-dimensional problems, neither in terms of storage requirement nor CPU burden.

#### 4.4.2 A progressive selection

To circumvent the issues identified above, we here use a bottom-to-top approach which achieves a forward stagewise regression by progressively revealing important basis functions. Introduced by Efron *et al.* (2004); Hastie *et al.* (2009), the Least Angle Regression Selection (LARS) technique relies on analytical solutions to speed-up computations and essentially follows the piecewise linear regularization path of the LASSO. In a nutshell, it

consists in selecting, from the *a priori* set  $\mathcal{J}_{\text{prior}}$ , the predictor (approximation function) which is most correlated with the current residual, move this predictor to the active set  $\mathcal{J}_{\text{post}}$ , compute the increment solution vector by minimizing the residual  $L^2$ -norm and follow the descent direction along the increment vector until a predictor from the inactive set becomes as correlated with the residual as those from the active set. The whole process is then repeated and allows to sequentially build the optimal subset of approximation functions by exploring the Pareto front defined by the competition between the two terms of the unconstrained formulation of the optimization problem of Eq. (23). One advantage of LARS over other techniques is that the whole potential dictionary is never stored nor used. A LARS approach in the UQ framework was also considered in Blatman & Sudret (2011).

We consider the following polynomial approximation  $\tilde{f}_\gamma$  of  $f_\gamma$ :

$$f_\gamma(\{\xi_i\}_{i \in \gamma}) \approx \tilde{f}_\gamma(\{\xi_i\}_{i \in \gamma}) := \sum_{\alpha, |\alpha| \leq \tilde{p}} c_{\gamma, \alpha} \psi_\alpha(\{\xi_i\}_{i \in \gamma}), \quad \psi_\alpha = \prod_{i \in \gamma} \psi_{\alpha_i}(\xi_i), \quad (25)$$

with  $\alpha = (\alpha_i, i \in \gamma)$ ,  $\alpha_i \in \{1, \dots, \tilde{p}\}$ . Interaction modes  $\{f_\gamma\}$  are then approximated in  $\mathbb{P}_{\tilde{p}}$ , the space of polynomials with maximum total degree  $\tilde{p}$ , by modes  $\{\tilde{f}_\gamma\}$  linear in their coefficients.

In the present framework, the HDMR approximation format naturally leads to *groups* of predictors which importance in describing the QoI  $u$  follows a similar trend. These groups are defined by the subsets  $\{\mathcal{J}_\gamma\}$  of predictors which belong to a given interaction mode  $f_\gamma$ ,  $\mathcal{J}_\gamma = \{\psi_\alpha(\{\xi_i\}_{i \in \gamma})\}$ , and are likely to be strongly correlated. For instance, if the QoI exhibits a strong dependence on a given dimension  $\xi_j$ , one then wants to incorporate the whole set of predictors  $\{\psi_\alpha(\{\xi_i\}_{i \in \gamma})\}$ ,  $\gamma : j \in \gamma$  without evaluating their relevance individually. One then looks for an approximation which is sparse at the level of groups of functions. Note that grouping predictors significantly alleviates the computational cost associated with the subset selection as further discussed in section 4.7.

It is important to stress that this approximation format is made only for the subset selection step and is independent of the format the QoI is finally approximated in. The selection of groups reduces to selection of interaction modes  $f_\gamma$  and leaves the possibility for using different formats between the subset selection and the coefficients evaluation steps: an interaction mode found to be dominant is incorporated to the active dictionary  $\mathcal{J}_{f, \text{post}}$ , independently of the way its contribution to the approximation of  $u$  is actually determined in the end. Indeed, since the LARS technique only applies to predictors *linear* in their coefficients, an approximation  $\tilde{f}_\gamma$  of the form (25) is suitable for the selection of the dominant groups. However, the final approximation  $\hat{f}_\gamma$  of the retained  $f_\gamma$  may rely on predictors *nonlinear* in their coefficients: the subset selection step only serves to determine which interaction modes will be considered in the *a posteriori* approximation basis, the ‘skeleton’  $\{f_\gamma : \gamma \in \mathcal{J}_{f, \text{post}}\}$ .

The selection is made using a modified LARS approach and the following optimization

problem is solved:

$$\mathbf{c} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}|}} \|\mathbf{u} - \Psi \tilde{\mathbf{c}}\|_2^2 + \tau \sum_{\gamma \in \mathcal{J}_f} \|\tilde{\mathbf{c}}_\gamma\|_{K_\gamma}, \quad (26)$$

with  $\tau > 0$  the regularization parameter and  $\|\cdot\|_{K_\gamma}$  a norm induced by a positive definite matrix  $K_\gamma$ . All predictors within a group  $\gamma$  are here weighted similarly so that we use a scaled identity matrix  $K_\gamma = \mathbf{I}_{|\mathcal{J}_\gamma|} / |\mathcal{J}_\gamma|$ ,  $\forall \gamma \in \mathcal{J}_f$ . The regularization term is a combination of  $L^2$ - and  $L^1$ - norms and penalizes the  $L^1$ -norm of the ‘group’ vector to promote a collective behavior: either a group is basically active (non-zero  $K_\gamma$ -norm) or inactive, essentially disregarding the detailed behavior within the group. This group LARS (gLARS) strategy was first proposed in Yuan & Lin (2006) and the algorithm presented in Xie & Zeng (2010) was modified to solve the optimization problem (26).

#### 4.5 Approximating a random variable

We now discuss the general methodology for approximating a random variable  $u(\boldsymbol{\xi})$  from a finite set of its realizations. An *a priori* choice of a representation format was first made based on knowledge one may have on the expected form of the QoI, section 4.3. This finite-dimensional space may still be large w.r.t. the amount of available information and was adjusted based on the data through the subset selection procedure, the *a posteriori* step, section 4.4. This has selected a set of groups, or interaction modes,  $\{f_\gamma\}_{\gamma \in \mathcal{J}_{f,\text{post}}}$  deemed to contribute most to the HDMR representation of the QoI. The actual approximation of  $u$  will rely on these selected groups but does not bear restriction on the linearity w.r.t. the coefficients so that different suitable formats, possibly nonlinear, can then be considered, possibly leading to fewer number of coefficients than at the *a posteriori* step. to further reduce the number of coefficients to estimate.

One possibility is to determine an approximation of  $\{f_\gamma, \gamma \in \mathcal{J}_{f,\text{post}}\}$  in the space  $\mathbb{P}_p$  of polynomials with maximum total degree  $p$  along each dimensions as considered in the last section:

$$\begin{aligned} f_\gamma(\{\xi_i\}_{i \in \gamma}) &\approx \hat{f}_\gamma(\{\xi_i\}_{i \in \gamma}) = \sum_{\boldsymbol{\alpha}, |\boldsymbol{\alpha}| \leq p} c_{\gamma, \boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\{\xi_i\}_{i \in \gamma}), & \psi_{\boldsymbol{\alpha}}(\{\xi_i\}_{i \in \gamma}) &= \prod_{i \in \gamma} \psi_{\alpha_i}(\xi_i), \\ \boldsymbol{\alpha} &= (\alpha_i, i \in \gamma), & \alpha_i &\in \{1, \dots, p\}, & 1 \leq |\gamma| \leq N_l^{(\text{PC})} &\leq \min(N_l, p). \end{aligned} \quad (27)$$

The cardinality associated with the approximation of  $f_\gamma$  at a given iteration level  $l = |\gamma|$  is  $|\mathcal{J}_\gamma| = p! / (l! (p-l)!)$  and usually provides an accurate approximation with a low number of coefficients for low dimensions  $|\gamma|$ .

When the dimension  $|\gamma|$  increases beyond a prescribed threshold  $N_l^{(\text{PC})}$ , the number of terms in  $\hat{f}_\gamma$  decreases and eventually degenerates for  $|\gamma| > p$ . For these higher interaction order modes, a low-rank canonical decomposition of the interaction modes is instead

considered:

$$f_{\gamma}(\{\xi_i\}_{i \in \gamma}) \approx \hat{f}_{\gamma}(\{\xi_i\}_{i \in \gamma}) = \sum_{r=1}^{n_r} \prod_{i \in \gamma} \sum_{\alpha=1}^p c_{\gamma, \alpha}^{r, i} \psi_{\alpha}(\xi_i), \quad N_l^{(\text{PC})} < |\gamma| \leq N_l \leq d. \quad (28)$$

The maximum number of modes at a given interaction level  $l$  is  $d! / ((d-l)! l!)$ . Relying on an approximation in  $\mathbb{P}_p$  for interaction modes of order  $|\gamma| \leq N_l^{(\text{PC})}$  and on low-rank approximation for higher interaction order modes, with maximum rank  $n_r$ , the total cardinality of this approximation format is bounded from above by

$$|\mathcal{J}_{\text{prior}}| = \sum_{l=0}^{N_l^{(\text{PC})}} \frac{d! p!}{(d-l)! (l!)^2 (p-l)!} + \sum_{l=N_l^{(\text{PC})}+1}^{N_l} \frac{d!}{(d-l)! l!} n_r l p. \quad (29)$$

We denote  $\mathcal{J}_{f, \text{eff}}$  the set of modes  $\{\hat{f}_{\gamma}\}_{\gamma \subseteq \mathcal{J}_{f, \text{post}}}$  finally considered for the approximation of  $u$  and  $\mathcal{J}_{\text{eff}}$  the set of associated predictors  $\{\psi_{\alpha}\}$ . The interaction modes are estimated sequentially. Once a new mode is evaluated, the whole approximation may be updated by reevaluating the coefficients of the predictors already evaluated of the current evaluation set  $\{\mathcal{J}_{\text{eff}}\}$ . Let  $\mathbf{z} = (z^{(1)} \dots z^{(N_q)})^T$  be the residual vector after basis functions  $\hat{f}_{\gamma}, \gamma \in \mathcal{J}_{f, \text{eff}}$  have been evaluated. The coefficients involved in the next mode  $\hat{f}_{\gamma}, \gamma \in \mathcal{J}_{f, \text{post}} \setminus \mathcal{J}_{f, \text{eff}}$  to be evaluated are then determined. If  $\gamma$  is such that  $|\gamma| \leq N_l^{(\text{PC})}$ , they are computed from the following system of equations<sup>1</sup>:

$$\begin{cases} \mathbf{c}_{\gamma, \cdot} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}_{\gamma}|}} \|\mathbf{z} - \Psi \tilde{\mathbf{c}}\|_2, & \forall i \in \gamma, \quad \gamma \subseteq \mathcal{J}_{f, \text{post}}, \quad |\gamma| \leq N_l^{(\text{PC})}, \end{cases} \quad (30)$$

with  $\mathbf{c}_{\gamma, \cdot} = (\mathbf{c}_{\gamma, \alpha_i}, i \in \gamma)^T$  and

$$\begin{aligned} z^{(q)} &= u^{(q)} - \sum_{\gamma' \subseteq \mathcal{J}_{f, \text{eff}} \setminus \gamma} \hat{f}_{\gamma'}(\{\xi_i^{(q)}\}_{i \in \gamma'}), & \mathbf{z} &= (z^{(1)} \dots z^{(N_q)})^T, \\ \Psi_{q\alpha} &= \psi_{\alpha}(\{\xi_i^{(q)}\}_{i \in \gamma}), & \Psi &= [\Psi_{q\alpha}]. \end{aligned} \quad (31)$$

To solve for the coefficients associated with predictors nonlinear in their coefficients, an Alternate Least Squares (ALS) approach is used, reformulating the nonlinear problem into a set of coupled linear equations:

$$\begin{cases} \mathbf{c}_{\gamma, \cdot}^{r, i} = \arg \min_{\tilde{\mathbf{c}} \in \mathbb{R}^p} \|\mathbf{z}_i - \Psi \tilde{\mathbf{c}}\|_2, & \forall i \in \gamma \subseteq \mathcal{J}_{f, \text{post}}, \quad N_l^{(\text{PC})} < |\gamma| \leq N_l, \end{cases} \quad (32)$$

<sup>1</sup> While not found necessary here, the solution of the least squares problem may be regularized by adding a generic term of the form  $\beta \|\mathbf{L} \tilde{\mathbf{c}}\|_2$ . A typical choice is  $L = I_{|\mathcal{J}_{\gamma}|}$  but one may also want to consider non-diagonal matrices  $L$ .

with  $\mathbf{c}_{\gamma,\cdot}^{r,i} = (\mathbf{c}_{\gamma,1}^{r,i} \dots \mathbf{c}_{\gamma,p}^{r,i})^T$  and

$$\begin{aligned} z_i^{(q)} &= u^{(q)} - \sum_{\gamma' \subseteq \mathcal{J}_{f,\text{eff}}} \widehat{f}_{\gamma'} \left( \left\{ \xi_i^{(q)} \right\}_{i \in \gamma'} \right) - \sum_{r'=1}^{r-1} \prod_{i' \in \gamma} \sum_{\alpha'=1}^p c_{\gamma,\alpha'}^{r',i'} \psi_{\alpha'} \left( \xi_{i'}^{(q)} \right), \\ \Psi_{q\alpha} &= \psi_{\alpha} \left( \xi_i^{(q)} \right) \prod_{i' \in \gamma, i' \neq i} \sum_{\alpha'=1}^p c_{\gamma,\alpha'}^{r,i'} \psi_{\alpha'} \left( \xi_{i'}^{(q)} \right), \quad \Psi = [\Psi_{q\alpha}]. \end{aligned} \quad (33)$$

This whole step is embedded in a loop over the modes  $f_{\gamma}, \gamma \in \mathcal{J}_{f,\text{post}}$  retained by the subset selection procedure. The cross-validation error ( $\text{CV}\varepsilon$ ) is estimated from  $\widehat{N}_q$  validation samples  $\left\{ \xi^{(\widehat{q})}, u^{(\widehat{q})} \right\}_{\widehat{q}=1}^{\widehat{N}_q}$  independent from the  $N_q$  samples of the training test. If the cross-validation error has increased over the last two loops, the approximation basis is likely to have become too large w.r.t. the available data and basis building iterations are stopped. The retained basis is then the one that has led to the lowest  $\text{CV}\varepsilon$ . On the other hand, if  $\text{CV}\varepsilon$  keeps decreasing, the next interaction mode as selected by the subset selection step is considered and added to the current active set  $\mathcal{J}_{f,\text{eff}}$  and the whole iteration is carried-out. Once the approximation is determined, the coefficients are updated with the same sequential technique using both the training and the validation points,  $N_q + \widehat{N}_q$ . The final approximation error  $\varepsilon$  is then estimated with the  $\widehat{N}_q$  test samples, Eq. (22). The global methodology is summarized in Algorithm 2.

#### 4.6 Robust estimation

An important concern when deriving a methodology is the robustness w.r.t. noise, in particular when dealing with experimental data as in the present work. We hence investigate this point and propose a more robust alternative to the methodology discussed so far.

To evaluate the approximation coefficients once an approximation basis is determined from the subset selection step, a standard approach is to minimize a norm between target observations and reconstructed approximation as done in the previous section, Eqs. (30) and (32): the approximation coefficients of a given mode  $\widehat{f}_{\gamma}$  are basically given by  $\mathbf{c}_{\gamma} = \arg \min_{\widetilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}_{\gamma}|}} \|\mathbf{z} - \Psi \widetilde{\mathbf{c}}\|_2$ , with  $\Psi \in \mathbb{R}^{N_q \times |\mathcal{J}_{\gamma}|}$  the matrix of the  $\gamma$ -group predictors evaluated in  $\left\{ \xi^{(q)} \right\}$  and  $\mathbf{z}$  the target residual vector. The solution to this least squares problem is equivalently obtained from

$$\{\mathbf{c}, \Delta \mathbf{z}\} = \arg \min_{\widetilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}_{\gamma}|}} \left\| \widetilde{\Delta} \mathbf{z} \right\|_F \quad \text{s.t.} \quad \mathbf{z} + \widetilde{\Delta} \mathbf{z} = \Psi \widetilde{\mathbf{c}}, \quad (34)$$

which minimizes the Frobenius norm of the residual vector. This implicitly assumes no error in the coordinates  $\left\{ \xi^{(q)} \right\}$  at which the target is evaluated. In practice, these coordinates are often not directly measurable and are inferred from observations via

---

**Algorithm 2** Sketch of the solution method for approximating a random variable  $u(\boldsymbol{\xi})$ 


---

- 1: Select an *a priori* basis in HDMR format. Choose  $p$ ,  $N_l$ ,  $N_l^{(\text{PC})}$ ,  $n_r$  and  $\tilde{p}$ . Initialize  $\mathbf{z} = \left(u^{(1)} \dots u^{(N_q)}\right)^T$ .
  - 2: **Subset selection step.** Solve the LASSO optimization problem with the gLARS algorithm  $\rightarrow$  sequence of *a posteriori* approximation bases indexed by  $s$  with ordered groups  $\mathcal{J}_{f,\text{post}} = \{\boldsymbol{\gamma}^{(s)}\}$ . Initialize  $s$  and  $\mathcal{J}_{f,\text{eff}}$ :  $s \leftarrow 0$ ,  $\mathcal{J}_{f,\text{eff}} \leftarrow \emptyset$ .
  - 3: **Solve the approximation problem:**
  - 4: **repeat**
  - 5:    $s \leftarrow s + 1$ .
  - 6:   Consider the next mode  $\hat{f}_{\boldsymbol{\gamma}^{(s)}}$  from the set  $\mathcal{J}_{f,\text{post}}$  selected in (2):  $\mathcal{J}_{f,\text{eff}} \leftarrow \mathcal{J}_{f,\text{eff}} \cup \boldsymbol{\gamma}^{(s)}$ .
  - 7:   Solve for the approximation coefficients  $\{\mathbf{c}_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma} \in \mathcal{J}_{f,\text{eff}}}$  by alternately solving for the coefficients of modes  $\{\hat{f}_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma} \in \mathcal{J}_{f,\text{eff}}}$ , Eqs. (30, 32). [**Update step**]
  - 8:   Estimate the cross-validation error  $\text{CV}\varepsilon$  and evaluate the current approximation  $\hat{\mathbf{u}} = \left(\hat{u}^{(1)} \dots \hat{u}^{(N_q)}\right)^T$ .
  - 9:   Update the residual  $\mathbf{z} \leftarrow \mathbf{u} - \hat{\mathbf{u}}$ .
  - 10: **until**  $\text{CV}\varepsilon$  has increased over the last two passes  $s$  and  $s - 1$ .
  - 11:  $\mathcal{J}_{f,\text{eff}} \leftarrow \mathcal{J}_{f,\text{eff}} \setminus \{\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s-1)}\}$ .
  - 12: Update the coefficients  $\{\mathbf{c}_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma} \in \mathcal{J}_{f,\text{eff}}}$  of the retained modes with the extended set of data  $\{\boldsymbol{\xi}^{(q)}, u^{(q)}\}_{q=1}^{N_q + \widehat{N}_q}$ . It finally yields  $\hat{u}(\boldsymbol{\xi})$  expressed in the basis  $\{\hat{f}_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma} \in \mathcal{J}_{f,\text{eff}}}$ .
- 

auxiliary inverse problems. This brings errors so that the actual coordinates vector is only estimated with an error  $\Delta \boldsymbol{\xi}^{(q)}$ . Since  $\Psi$  depends on  $\boldsymbol{\xi}$ , an error predictor matrix  $\Delta \Psi(\boldsymbol{\xi}, \Delta \boldsymbol{\xi}) := \Psi(\boldsymbol{\xi} + \Delta \boldsymbol{\xi}) - \Psi(\boldsymbol{\xi})$  arises and the estimation problem (34) then rewrites

$$\{\mathbf{c}, \Delta \Psi, \Delta \mathbf{z}\} = \arg \min_{\widetilde{\mathbf{c}} \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{\gamma}}|}} \left\| \widetilde{\Delta \Psi} \widetilde{\Delta \mathbf{z}} \right\|_F \quad \text{s.t.} \quad \mathbf{z} + \widetilde{\Delta \mathbf{z}} = \left( \Psi + \widetilde{\Delta \Psi} \right) \widetilde{\mathbf{c}}. \quad (35)$$

The realizations of the error in the data  $\{\Delta \boldsymbol{\xi}^{(q)}(\theta), \Delta z^{(q)}(\theta)\}$  are modeled to follow the distribution of zero-mean *iid* variables. Further, predictors may be correlated:

$$\mathcal{E}_{\theta} \left[ \left( \Delta \psi_{\boldsymbol{\alpha}} - \mathcal{E}_{\theta} [\Delta \psi_{\boldsymbol{\alpha}}] \right) \left( \Delta \psi_{\boldsymbol{\alpha}'} - \mathcal{E}_{\theta} [\Delta \psi_{\boldsymbol{\alpha}'}] \right) \right] \neq 0, \quad (36)$$

with  $\Delta \psi_{\boldsymbol{\alpha}} = \Delta \psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}^{(q)}, \Delta \boldsymbol{\xi}^{(q)})$ . A general approach to solve the weighted Total Least Squares (wTLS) problem of Eq. (35) consists in the minimization of the usual weighted residual sum of squares  $\rho^2$ , Markovsky & van Huffel (2007):

$$\rho^2 := \text{vec}(\Delta X)^T \Lambda^{-1} \text{vec}(\Delta X), \quad \Delta X := (\Delta \Psi \Delta \mathbf{z})^T, \quad X := (\Psi \mathbf{z})^T, \quad (37)$$

where the ‘vec’ operator unfolds a generic  $m \times n$  matrix into a  $mn$  vector and  $X$  is the data matrix. The covariance matrix for  $X$ ,  $\Lambda := \left\langle \text{vec} \left( X - \langle X \rangle_{N_q} \right) \text{vec} \left( X - \langle X \rangle_{N_q} \right)^T \right\rangle_{N_q}$ , is evaluated and the minimization problem (35) is solved using the ALS-based algorithm



proposed in Wentzell *et al.* (1997).

As will be shown in the numerical experiments examples, section 5.1.4, the present total least squares formulation allows to improve the approximation quality from noisy data. However, the solution process is significantly slower than that using the standard least squares. The present paper is based on the assumption that the critical part of the whole solution chain of determining a good approximation of the QoI is the data acquisition and that the cost of the post-processing part is not the main issue. However, if the computational cost is a strong concern, the robust approach should only be employed when noise in the data cannot be neglected with regard to the targeted accuracy of the approximation.

**Remark 2** *When a large amount  $N_q$  of experimental information is available, the data matrix  $X \in \mathbb{R}^{(|\mathcal{J}_\gamma|+1) \times N_q}$  can be large. The resulting correlation matrix  $\Lambda$  then has potentially very large dimensions. However, since the noise is assumed independent from one sample to another,  $\Lambda$  has a block diagonal structure. Further, it is a symmetric definite positive matrix, allowing for additional reduction of the storage requirement. The structure of  $\Lambda$  is then exploited in solving the weighted total least squares problem above through sparse storage and operations.*

#### 4.7 Numerical complexity

While the primary motivation for this work is to determine an accurate representation of a random quantity from a small set of its realizations, it is desirable that the solution method remains computationally tractable. As seen above, the algorithm for approximating a random variable is essentially two fold.

The selection process essentially consists in sequentially building a subset, section 4.4. Each step of the sequence involves solving a least squares problem of growing size and finding the basis function, or group of functions, within the *a priori* set  $\mathcal{J}_{\text{prior}}$  most correlated with the current residual. The matrix of the least squares problem is  $\Psi \in \mathbb{R}^{N_q \times |\mathcal{J}_{\text{post}}|}$ , with  $|\mathcal{J}_{\text{post}}|$  the cardinality of the current set of selected basis functions. The least squares problem is solved *via* a QR decomposition of  $\Psi$  in  $\mathcal{O}(N_q |\mathcal{J}_{\text{post}}|^2)$  operations. The iterative selection process is carried-out with a growing active set  $\mathcal{J}_{\text{post}}$  until the problem becomes ill-posed, *i.e.*, until  $|\mathcal{J}_{\text{post}}|$  is about  $N_q$ . We use grouped LARS and denote  $\overline{|\mathcal{J}_{\gamma, \text{post}}|}$  the average cardinality of the retained group predictors, *i.e.*, the average number of basis functions in the group added to the active set. The subset selection process retains  $n_f$  groups of variables so that the total cost associated with the least squares step of the subset selection is

$$\mathcal{J}_{\text{LS}} = \mathcal{O} \left( \sum_{s=1}^{n_f} N_q \left( \overline{|\mathcal{J}_{\gamma, \text{post}}|} s \right)^2 \right). \quad (38)$$

As groups of predictors are moved to the active set, the size of the remaining *a priori* set decreases,  $|\mathcal{J}_{\text{prior}}|^{(\text{current})} \simeq |\mathcal{J}_{\text{prior}}| - s \overline{|\mathcal{J}_{\gamma, \text{post}}|}$ . The cost associated with the evaluation of

the correlation for each predictor in the inactive set is then:

$$\mathcal{J}_{\text{correl}} = \mathcal{O} \left( \sum_{s=1}^{n_f} N_q \left( |\mathcal{J}_{\text{prior}}| - s \overline{|\mathcal{J}_{\gamma, \text{post}}|} \right) \right) \propto N_q. \quad (39)$$

In practice, the cost associated with the evaluation of the correlation of the predictors in the inactive set with the current residual dominates so that the whole cost of the subset selection finally approximates as

$$\mathcal{J}_{\text{subsel}} = \mathcal{J}_{\text{LS}} + \mathcal{J}_{\text{correl}} \simeq \mathcal{O} \left( N_q |\mathcal{J}_{\text{prior}}| n_f - N_q \overline{|\mathcal{J}_{\gamma, \text{post}}|} \frac{n_f (n_f + 1)}{2} \right). \quad (40)$$

The second step of the solution method deals with the evaluation of the approximation coefficients, section 4.5. The cost associated with evaluating the coefficients of a  $l$ -th interaction order mode,  $1 \leq l \leq N_l^{(\text{PC})}$ , encompasses the matrix  $\Psi$  assembly cost  $\mathcal{O}(N_q p! / (l! (p-l)!))$  and the least squares solution  $\mathcal{O}(N_q (p! / (l! (p-l)!))^2)$ . Since modes  $\{\hat{f}_\gamma\}_{\gamma \in \mathcal{J}_{f, \text{eff}}}$  already evaluated may be updated once an additional one from the selected set is considered, the total cost is the sum of an arithmetic sequence. Its exact formulation depends on the selected set and is difficult to derive in closed-form. As a simple example, updating all coefficients for each new mode  $\hat{f}_\gamma$  considered, neglecting the cost associated with first-order interaction modes and assuming only second-order interaction modes are retained in the *a posteriori* set, an upper bound for the cost writes

$$\mathcal{J}_{\text{coef}} \leq \sum_{s=1}^{|\mathcal{J}_{f, \text{eff}}|} \left[ \mathcal{O}(s N_q p^l) + \mathcal{O}(s N_q p^{2l}) \right], \quad \text{with } l = 2, \quad (41)$$

where  $|\mathcal{J}_{f, \text{eff}}|$  is the number of groups finally retained for the approximation by the CV test, see Algorithm 2. A quantitative discussion of the numerical cost is given in section 5.1.5 with an illustrative example.

#### 4.8 Statistics and sensitivity analysis

Once an approximation of the random variable  $u$  is obtained, it is easy to estimate its first statistical moments. The methodology described in sections 4.3–4.5 can be applied ‘as is’, without a connection with approximating a random process. It is then useful to discuss how to determine basic statistics of the approximated random variable. From the HDMR format properties, the estimated mean is simply given by the first term of the decomposition:  $\langle u \rangle_{L^2(\Xi, \mu_\Xi)} \simeq f_\emptyset$ .

Thanks to the orthogonality property of the modes  $\{f_\gamma\}$ , the variance  $\text{Var}(u) := \left\langle \left( u - \langle u \rangle_{L^2(\Xi, \mu_\Xi)} \right)^2 \right\rangle_{L^2(\Xi, \mu_\Xi)}$  approximates as the sum of the variance of the individual interaction modes:

$$\begin{aligned}
\text{Var}(u) &\simeq \sum_{\gamma \in \mathcal{J}_{f,\text{eff}} \setminus \emptyset} \text{Var}(\hat{f}_\gamma), \\
&= \sum_{\substack{\gamma \in \mathcal{J}_{f,\text{eff}} \setminus \emptyset \\ |\gamma| \leq N_l^{(\text{PC})}}} \sum_{\alpha, |\alpha| \leq p} c_{\gamma,\alpha}^2 \|\psi_\alpha\|_{L^2(\Xi, \mu_\Xi)}^2 + \sum_{\substack{\gamma \in \mathcal{J}_{f,\text{eff}} \setminus \emptyset \\ N_l^{(\text{PC})} < |\gamma| \leq N_l}} \sum_{r,r'=1}^{n_r(\gamma)} \prod_{i \in \gamma} \sum_{\alpha=1}^p c_{\gamma,\alpha}^{r,i} c_{\gamma,\alpha}^{r',i} \|\psi_\alpha\|_{L^2(\Xi, \mu_\Xi)}^2, \quad (42)
\end{aligned}$$

where use was made of the orthogonality of the Hilbertian basis  $\{\psi_\alpha\}$ .

Other standard statistical quantities are the sensitivity indices  $\{S_\gamma\} := \text{Var}(\hat{f}_\gamma) / \text{Var}(\hat{u})$  which essentially represent the relative part of the variance of the QoI due to the interaction of a given set of input random variables only, Sobol (1993); Homma & Saltelli (1996). From Eq. (42), it immediately follows that  $\sum_{\gamma \subseteq \{1, \dots, d\} \setminus \emptyset} S_\gamma = 1$  and the explicit expression of the sensitivity indices is straightforward to derive from the HDMR format. In practice, it is often more useful to assess the influence of a given input onto the variance of the QoI with the total sensitivity indices  $\{S_{T,i}\}_{i=1}^d$ :

$$S_{T,i} := \frac{\sum_{\gamma \subseteq \{1, \dots, d\} \setminus \emptyset: i \in \gamma} \text{Var}(\hat{f}_\gamma)}{\text{Var}(\hat{u})}, \quad 1 \leq i \leq d. \quad (43)$$

Again, using Eq. (42), this quantity is straightforward to estimate once the approximation of  $u(\boldsymbol{\xi})$  is available.

## 5 Numerical experiments

The methodology developed in the previous sections is now demonstrated on a set of examples. Different aspects of the global solution method are illustrated on a 1-D stochastic diffusion equation. A more computationally involved example is next considered with a Shallow Water problem with multiple sources of uncertainty.

### 5.1 Approximating a random variable

We consider the 1-D stochastic diffusion equation presented in section 4.3, briefly recalled here for sake of convenience:

$$\nabla_x (\nu(x, \boldsymbol{\xi}) \nabla_x u(x, \boldsymbol{\xi})) = F(x, \boldsymbol{\xi}), \quad u(x_-, \boldsymbol{\xi}) = u_-, \quad u(x_+, \boldsymbol{\xi}) = u_+. \quad (44)$$

The stochastic approximation basis relies on a HDMR format with a maximum interaction order  $N_l = 3$  and 1-D Legendre polynomials  $\{\psi_\alpha\}_{\alpha=1}^p$  of maximum degree  $p = 8$ .

In this section, the focus is on approximating a purely random quantity, *i.e.*, disregarding its spatial dependence. We then rely on samples of the solution  $u(x, \boldsymbol{\xi})$  taken at a given spatial location  $x^*$ :  $\{u^{(q)} := u(x^*, \boldsymbol{\xi}^{(q)})\}_{q=1}^{N_q}$ .

### 5.1.1 Influence of the number of samples

We first focus on the achieved accuracy in the approximation with a given budget  $N_q + \widehat{N}_q$  samples. The number of test points  $\widetilde{N}_q$  to estimate the approximation error  $\varepsilon$ , Eq. (22), is chosen sufficiently large so that  $\varepsilon$  is well estimated,  $\widetilde{N}_q = 10,000$ . In Fig. 3, the performance of the present gLARS-ALS methodology is compared with both a plain HDMR approximation, *i.e.*, with no subset selection hence considering the whole *a priori* approximation basis, and a Polynomial Chaos (PC) approximation with a sparse grid technique. The Smolyak scheme associated with a Gauss-Patterson quadrature rule is used as the sparse grid, with varying number of points in the 1-D quadrature rule and varying levels. The dimensionality of the stochastic space is  $d = 8$ .

The sparse grid is seen to require a large number of samples to reach a given approximation accuracy.<sup>2</sup> The HDMR-format approximation, with various interaction orders  $N_l$ , provides a better performance than PC/Smolyak but still requires more points to reach a given accuracy than the present methodology which performs significantly better in approximating the QoI from a given dataset. The present gLARS-ALS approximation error is also seen to be smooth and monotonic when the amount of information varies. When  $N_q$  is large enough, the subset selection step becomes useless as all  $|\mathcal{J}_{\text{prior}}|$  terms of the proposal basis can be estimated from the large amount of information and the present gLARS-ALS performance should then be similar as that of the HDMR. This is indeed what is observed for  $N_q \gtrsim 5000$ . Note that the benefit of a subset selection step in terms of accuracy improvement increases with the dimension  $d$  as the size  $|\mathcal{J}_{\text{prior}}|$  of the potential dictionary then grows.

### 5.1.2 Influence of the stochastic dimension

The approximation accuracy of the present method is now studied when the dimension of the stochastic space varies. The same problem as above is considered but with various truncation orders of the source  $F$  and the diffusion coefficient  $\nu$  definitions, see Eqs. (20). The solution of the diffusion problem (44) is of dimension  $d = d_F + d_\nu$  and the dimensions  $d_F$  and  $d_\nu$  are varied together,  $d_F = d_\nu$ . The resulting approximation error is plotted in Fig. 4 for different  $d$  when the number of available samples varies. From  $d = 8$  to  $d = 40$ , the required number of points for a given accuracy is seen to increase significantly, between a 2- and a 10-fold factor. However this is much milder than the increase in the potential

---

<sup>2</sup> Note that the plain Smolyak scheme is used here, which does not exploit anisotropy in the response surface. More sophisticated Smolyak-based approximations have been developed, see Nobile *et al.* (2007), and are expected to provide better results.

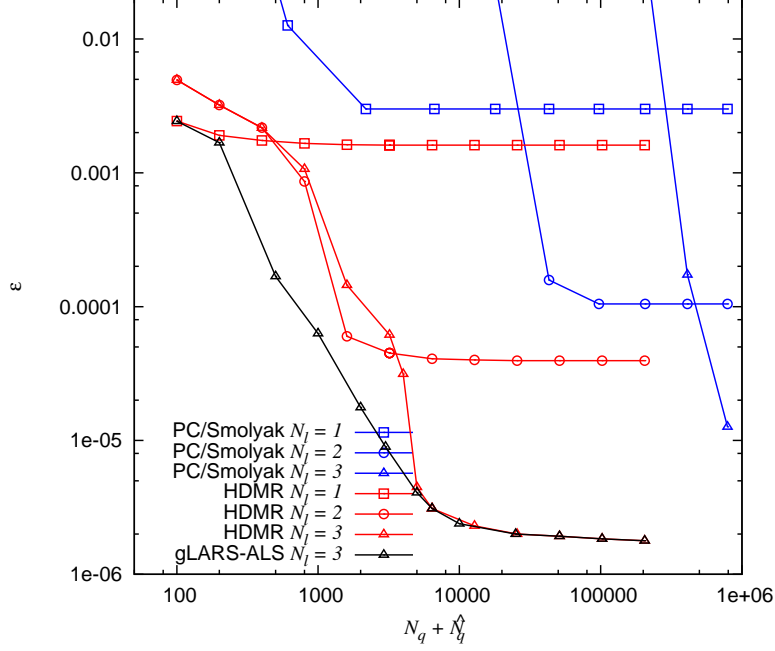


Fig. 3. Convergence of the approximation with the number of samples  $N_q + \widehat{N}_q$ . Different approximation methods are compared: plain HDMR, PC/Smolyak scheme sparse grid spectral decomposition and the present gLARS-ALS. The convergence is plotted in terms of  $\varepsilon$ .  $d = 8$ ,  $p = 8$ ,  $N_l = 3$ ,  $N_l^{(\text{PC})} = 3$ .

approximation basis cardinality, *i.e.*, if not subset selection was done, as  $|\mathcal{J}_{\text{prior}}|$  shifts from 10,565 ( $d = 8$ ) to  $1.7 \times 10^6$  ( $d = 40$ ), demonstrating the efficiency of the subset selection step which activates only a small fraction of the dictionary. When  $d$  further increases from 40 to 100, the performance remains essentially the same with hardly any loss of accuracy for a fixed  $N_q$ : the solution method is able to capture the low-dimensional manifold onto which the solution essentially lies and an increase in the size of the solution space hardly affects the number of samples it requires. This capability is a crucial feature when available data is scarce and the solution space is very large. As an illustration, when  $d = 100$ , and with the parameters retained, the potential cardinality of the approximation basis is about  $27 \times 10^6$  while the number of available samples is  $\mathcal{O}(100 - 10,000)$ . It clearly illustrates the pivotal importance of the subset selection step. Note that if one substitutes a Polynomial Chaos approximation to the present HDMR format, about  $352 \times 10^9$  terms need be evaluated with the present settings, a clearly daunting task.

For sake of completeness, the approximation given by a CP-format, Eq. (13), is also considered. The univariate functions  $\{f_{i,r}\}$  are approximated with the same polynomial approximation as in the present gLARS-ALS approach and a Tikhonov-based regularized ALS technique is used to determine each  $f_{i,r}$  in turn given the others. Upon convergence, the next set of modes  $\{f_{1,r+1}, \dots, f_{d,r+1}\}$  is evaluated until a maximum rank  $n_r$  set by cross-validation. At each rank  $r$ , the best approximation, as estimated by cross-validation, is retained from a set of initial conditions and regularization parameter values. As can be appreciated from Fig. 4, the number of samples required for a given approximation error

is significantly larger than with the present gLARS-HDMR method.

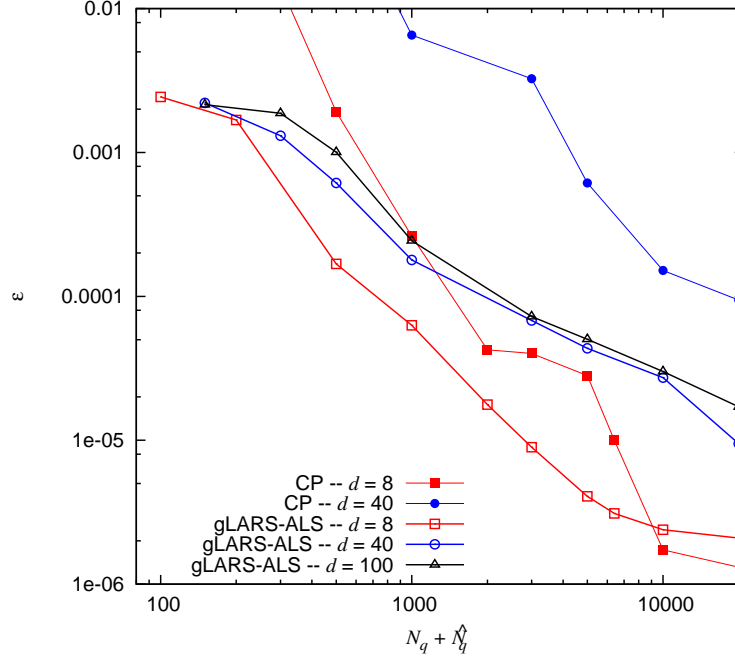


Fig. 4. Convergence of the approximation with the number of samples  $N_q + \widehat{N}_q$  and for different dimensionality of the QoI. The present gLARS-ALS approach is compared with a CANDECOMP-PARAFAC-type technique (labeled ‘CP’).

### 5.1.3 Subset selection

To further illustrate the subset selection step, the set of second and third order interaction retained modes  $\{f_\gamma\}_{\gamma \in \mathcal{J}_{f,\text{post}}}$  are plotted in Fig. 5 in the  $d = 40$  case. Each bullet represents one of the  $d$  stochastic dimensions and each line connects two (2-nd order, left plot) or three (3-rd order, right plot) dimensions, denoting a retained mode. The first  $d_F = 20$  of the 40 dimensions are associated with the source term  $F$  in the stochastic equation and are represented as the solid bullets of the first two quadrants,  $d \in [1, 20]$ . The other  $d_\nu = 20$  dimensions are associated with the uncertain diffusion coefficient  $\nu$  and are plotted as open bullets in the 3-rd and 4-th quadrants,  $d \in [21, 40]$ . The dimensions introduced by these two quantities are sorted with the associated magnitude of the eigenvalues  $\sigma_F$  and  $\sigma_\nu$  of their kernel, see Eqs. (21), which decreases along the counter-clockwise direction. Hence, the norm of the eigenvalues of the kernel associated with  $F$  decreases when one goes counter-clockwise from the first to the second quadrant. Likewise, the norm of the eigenvalues associated with dimensions introduced by  $\nu$  decreases from the third to the fourth quadrant. Dominant dimensions of the stochastic space for the output  $u$  approximation are thus expected to lie at the beginning of the first and/or third quadrant.

From the plot of second order modes (left), the subset selection process is seen to retain interaction modes mainly associated with dominant eigenvalues of both  $F$  and  $\nu$ : they

mainly link bullets from the first (dominant) dimensions associated with  $F$  to the first (dominant) dimensions associated with  $\nu$ , as one might expect. Further, modes associated with two dimensions both introduced by  $\nu$  are seen to be selected while two dimensions both associated with  $F$  are rarely connected: the subset selection procedure is able to capture the nonlinearity associated with  $\nu$  in the QoI and retains corresponding interaction modes. Indeed, note from Eq. (44) that the source term  $F$  interacts linearly with the solution  $u$  while the diffusion coefficient is nonlinearly coupled with  $u$  and hence, interaction modes between two dimensions introduced by  $F$  do not contribute to the approximation. The third order modes (right plot) also illustrate the nonlinearity associated with  $\nu$ : the retained modes either connect dimensions associated with  $\nu$  only or with one  $F$ -related and two  $\nu$ -related dimensions. Again, no two dimensions of  $F$  are connected, consistently with the linear dependence of  $u$  with  $F$ . These results illustrate the effectiveness of the procedure to unveil the dominant dependence structure and to discard unnecessary approximation basis functions.

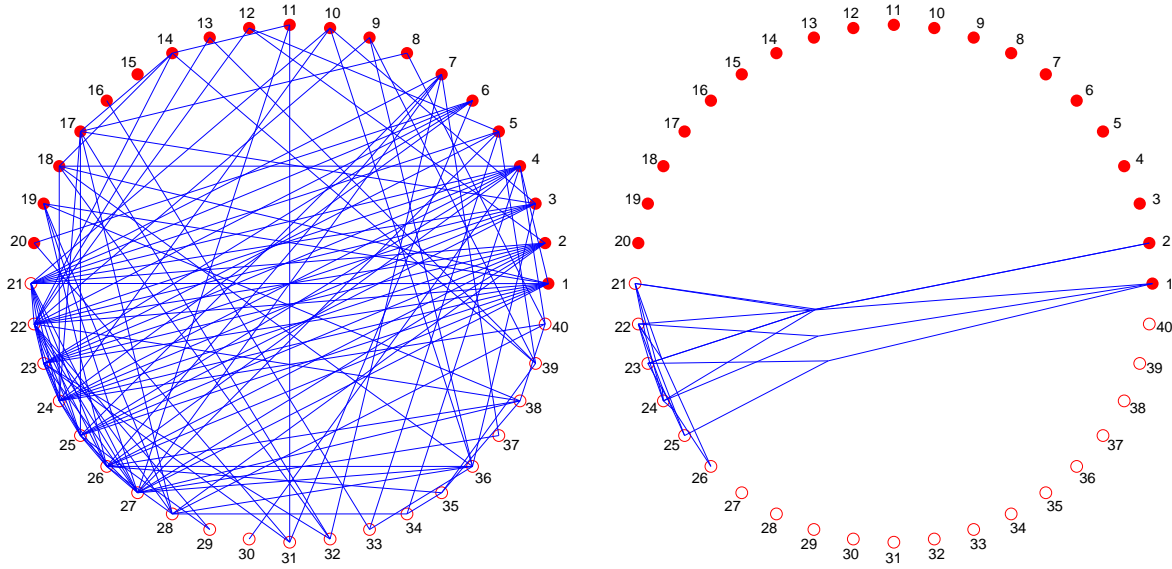


Fig. 5. Graphical representation of the interaction modes retained by the subset selection procedure. Left: second order modes are plotted as a line linking two dimensions (bullets). Right: third order modes are represented as 3-branch stars and connect three dimensions.

#### 5.1.4 Robustness

The robustness of the approximation against measurement noise is now investigated. The dataset  $\{\xi^{(q)}, u^{(q)}\}_{q=1}^{N_q}$  is corrupted with noise, mimicking variability introduced in practice in measurements and coordinates estimation. Denoting the nominal value with a star as superscript, noise in the coordinates is modeled as

$$\xi^{(q)} = \xi^{(q)*} + s \zeta^{(q)}, \quad \forall 1 \leq q \leq N_q, \quad (45)$$

with  $s > 0$ . The noise is modeled as an additive  $d$ -dimensional, zero-centered, unit variance, Gaussian random vector  $\zeta$  biased so that  $\xi^{(q)} \in [-1, 1]^d, \forall q$ . It is independent from



one sample  $q$  to another. Without loss of generality, measurements are here modeled as being corrupted with a multiplicative noise:  $u^{(q)} = u^{(q)*} (1 + s_u \zeta_u^{(q)})$ , with  $s_u = 0.2$  and  $\zeta_u \sim \mathcal{N}(0, 1)$ .

The evolution of the approximation accuracy when the noise intensity  $s$  in the coordinates varies is plotted in Fig. 6 in terms of error estimation  $\varepsilon$ . We compare gLARS-ALS using standard least squares (LS) with its ‘robust’ counterpart relying on weighted total least squares (wTLS) as discussed in section 4.6.

When the noise intensity increases, the error exponentially increases, quickly deteriorating the quality of the approximation with a noise standard deviation here as low as  $s = 3 \times 10^{-5}$ . When the noise is strong (low SNR), both the LS and the wTLS methods achieve poor accuracy. However, if the dataset is only mildly corrupted with noise, the wTLS approach is seen to achieve a significantly better accuracy than the standard least squares. While it is useful only on a range of SNR and somehow computationally costly, this feature is deemed important for a successful solution method in an experimental context where noise is naturally present.

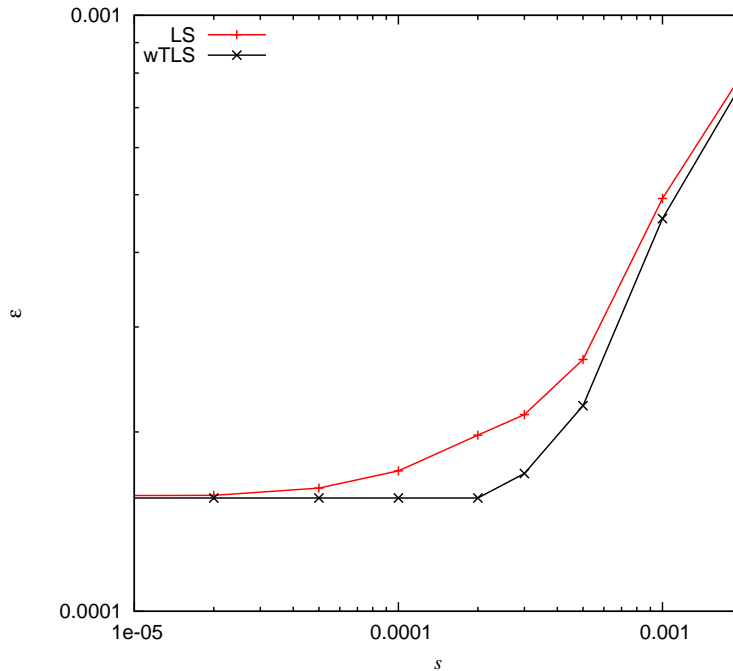


Fig. 6. Robustness of the approximation w.r.t. noise in the data: approximation error  $\varepsilon$  from the standard least squares (LS) and weighted total least squares (wTLS).  $d = 5$ ,  $N_q = 500$ .

### 5.1.5 Scaling of the solution method

In this section, the numerical complexity associated with the different steps of the solution method is illustrated in terms of computational time. Numerical experiments are carried-out with varying number of samples  $N_q$  and solution space dimensions  $d$ . When one is varying, the other remains constant. The nominal parameters are  $d = 40$  (dimension of the stochastic space),  $p = 6$  (maximum total order of the Legendre polynomials),

$N_q = 1000$  (number of samples),  $N_l = 3$  (maximum interaction order of the truncated HDMR approximation),  $\tilde{p} = 5$  (maximum total polynomial order in the subset selection step).

Numerical results are gathered in Fig. 7. The asymptotic behavior of the number  $n_f$  of required subset selection iterations as introduced in section 4.7 might be different according to which limit is considered. For the present stochastic diffusion problem, first and second interaction order modes tend to be selected first. Assuming the active set  $\mathcal{J}_{f,\text{post}}$  is dominated by first and second interaction order modes, it can easily be shown that the number of retained groups then satisfies

$$n_f \leq 1 + n_{f_1} + \min \left[ \frac{d(d-1)}{2}, 2 \frac{N_q - n_{f_1} \tilde{p}}{\tilde{p}(\tilde{p}+1)} \right], \quad n_{f_1} \leq \min \left[ d, \frac{N_q}{\tilde{p}} \right]. \quad (46)$$

In the present example, second order interaction groups dominate the retained set so that the number of retained groups tends to scale as  $n_f \propto N_q/\tilde{p}^2$ . From Eq. (40) and for the present nominal parameters, it results in the following limit behavior for the subset selection step:

$$\begin{aligned} \lim_{d \rightarrow +\infty} \mathcal{J}_{\text{subsel}} &\propto N_q^2 |\mathcal{J}_{\text{prior}}|/\tilde{p}^2 & \longrightarrow \text{here : } &\propto N_q^2 d^{N_l} \tilde{p}^{N_l-2}, \\ \lim_{N_q \rightarrow +\infty} \mathcal{J}_{\text{subsel}} &\propto N_q^2 |\mathcal{J}_{\text{prior}}|/\tilde{p}^2 & \longrightarrow \text{here : } &\propto N_q^2 d^{N_l} \tilde{p}^{N_l-2}. \end{aligned} \quad (47)$$

Similarly, the cost associated with the coefficients evaluation is considered. The number of interaction modes  $\mathcal{J}_{f,\text{eff}}$  effectively varies between 1 and  $\mathcal{O}(N_q/p^2)$  along the solution procedure, and, since the cost associated with solving the least squares problem dominates that of the matrix assembly, the cost of their evaluation finally simplifies in  $\mathcal{J}_{\text{coef}} \propto \mathcal{O}(N_q^2 p)$  or  $\mathcal{J}_{\text{coef}} \propto \mathcal{O}(N_q^3/p)$  depending on whether the coefficients are updated whenever an additional group is considered or not, see section 4.5 and second bullet of step (3) in Algorithm 2. In the present regime, the cost is found not to depend on  $d$ .

These asymptotic behaviors are consistent with the numerical experiments as can be appreciated from Fig. 7. The coefficients are here updated whenever a new mode from the selected set is considered, hence  $\mathcal{J}_{\text{coef}} \propto \mathcal{O}(N_q^3/p)$ . It is seen that the subset selection step scales less favorably than the coefficients evaluation step with the dimensionality of the random variable. This stresses the benefit of a carefully chosen *a priori* approximation basis to reduce as much as possible the cardinality  $|\mathcal{J}_{\text{prior}}|$  of the potential dictionary the final set is determined from by the subset selection step.

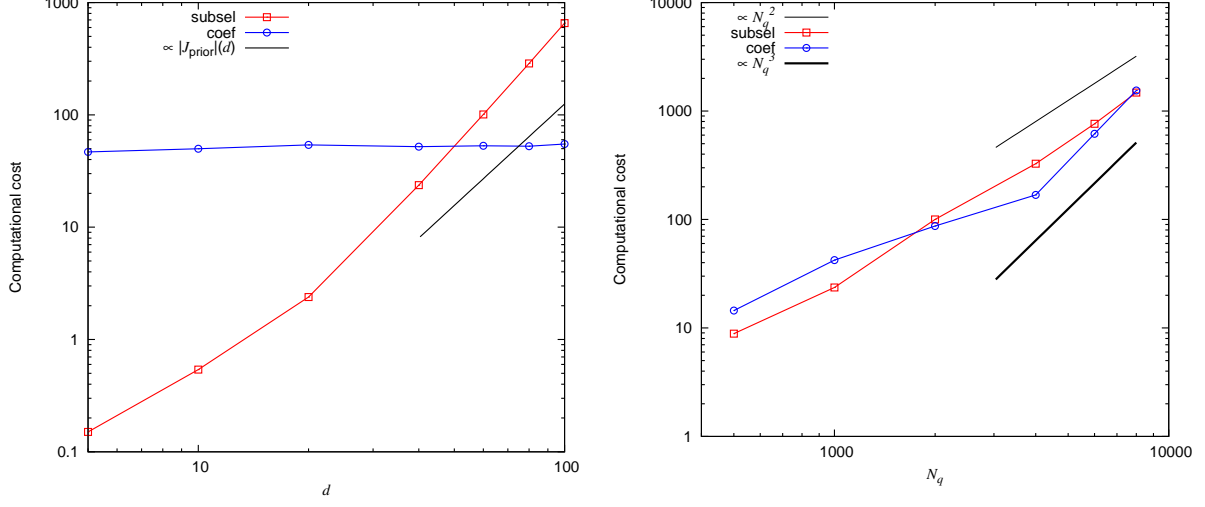


Fig. 7. Numerical cost of the subset selection and coefficients evaluation steps as a function the stochastic dimension  $d$  and size of the dataset  $N_q$ . Approximation coefficients are fully updated for each new mode. Nominal parameters are  $d = 40$ ,  $p = 6$ ,  $N_q = 1000$ ,  $N_l = 3$ ,  $\tilde{p} = 5$ ,  $N_l^{(\text{PC})} = 3$ .

## 5.2 Stochastic process approximation

We now consider the approximation of the space-dependent random solution  $u(x, \xi)$  under the form (17) using Algorithm 1. The approximation obtained from different number of samples  $\{x^{(q)}, \xi^{(q)}, u^{(q)}\}$  is compared with the Karhunen-Loève modes, computed from a full knowledge of the QoI, which henceforth constitute the reference solution. The simulation relies on the following parameters:  $|\mathcal{J}_x| = 32$ ,  $p = 10$ ,  $d = 6$ ,  $N_l = 3$ ,  $N_l^{(\text{PC})} = 3$ . The potential approximation basis cardinality is about  $|\mathcal{J}_x| |\mathcal{J}_{\text{prior}}| \simeq 10^5$ .

Fig. 8 shows the first and second spatial modes,  $w_1(x)$  and  $w_2(x)$  for different sizes of the dataset,  $N_q = 1000, 3000, 9000$  and  $26,000$ . The mean mode  $w_0(x)$  is virtually indistinguishable from the reference solution mean mode for any of the dataset sizes and is not plotted. On the left plot ( $w_1(x)$ ), it is seen that the approximation is decent, even with as low as  $N_q = 1000$  samples. For  $N_q = 3000$ , the approximation is good. This  $(1 + d) = 7$ -dimensional case corresponds to  $N_q^{1/(1+6)} \simeq 3.1$  samples per solution space dimension only and about  $N_q / (|\mathcal{J}_x| |\mathcal{J}_{\text{prior}}|) \simeq 3\%$  of the potentially required information.

For approximating the second spatial mode (Fig. 8, right plot), more points are needed to reach a good accuracy but  $N_q = 26,000$  is seen to already deliver a good performance. Quantitative approximation error results are gathered in Table 2 for various separation ranks  $N$  and number of samples  $N_q$ .

The satisfactory performance of the present method can be understood from the upper part of the Karhunen-Loève approximation (normalized) spectrum plotted in Fig. 9. The norm of the eigenvalues decays quickly so that the first two modes contribute more than 90 % of the QoI  $L^2$ -norm, showing that this problem efficiently lends itself to the present separation of variables-based methodology.

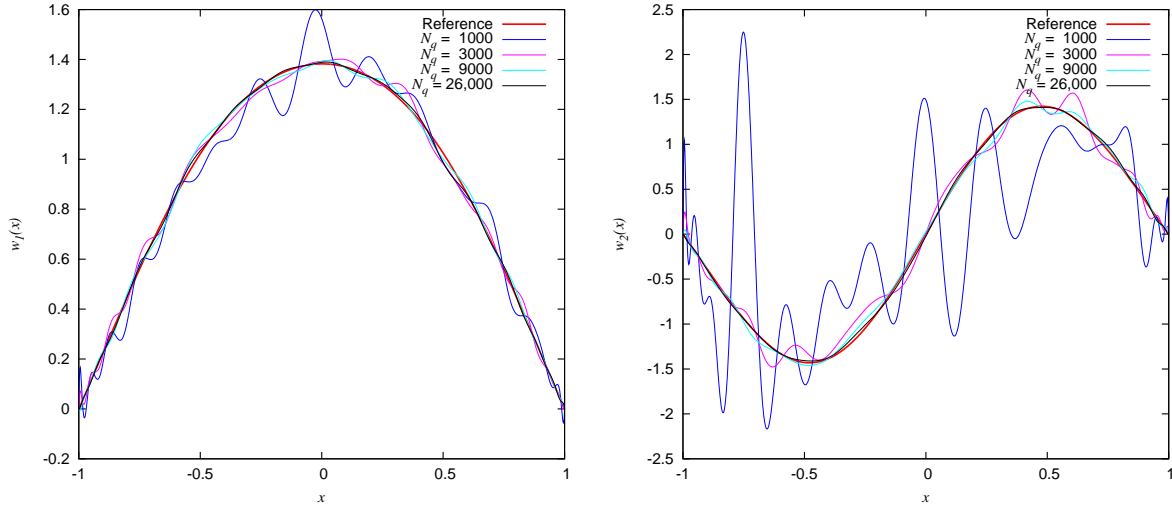


Fig. 8. First ( $w_1(x)$ , left) and second ( $w_2(x)$ , right) spatial approximation modes of the stochastic diffusion solution. The reference (Karhunen-Loève) solution is plotted for comparison (thick line).

$N_q \setminus N$	0	1	2
1000	$5.5 \times 10^{-3}$	$7.4 \times 10^{-4}$	$7.4 \times 10^{-4}$
3000	$5.5 \times 10^{-3}$	$4.2 \times 10^{-4}$	$2.7 \times 10^{-4}$
9000	$5.5 \times 10^{-3}$	$3.1 \times 10^{-4}$	$1.0 \times 10^{-4}$
26,000	$5.4 \times 10^{-3}$	$2.8 \times 10^{-4}$	$6.2 \times 10^{-5}$

Table 2

Evolution of the approximation error  $\varepsilon$  with the decomposition rank  $N$  and the number of samples  $N_q$ .

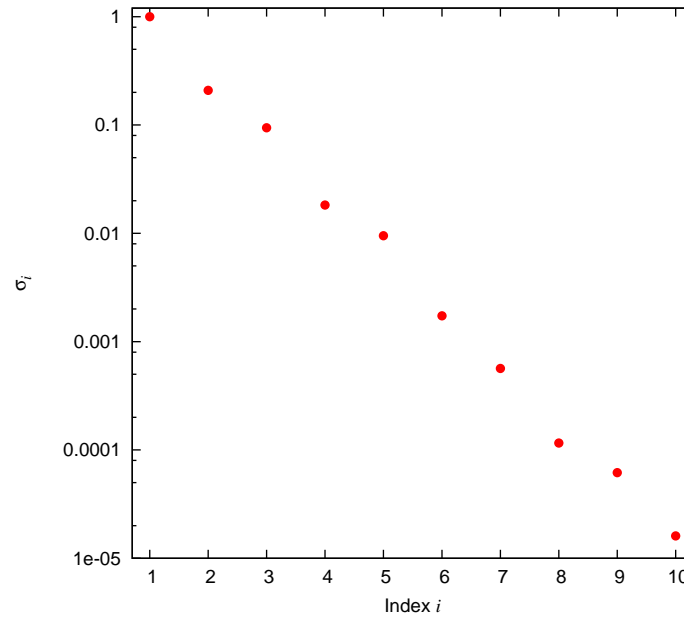


Fig. 9. Normalized upper spectrum of the Karhunen-Loève approximation.

### 5.3 A Shallow Water flow example

The methodology is now applied to the approximation of the stochastic solution of a Shallow Water flow simulation with multiple sources of uncertainty. It is a simple model for the simulation of wave propagation on the ocean surface. Waves are here produced by the sudden displacement of the sea bottom at a given magnitude, extension and location, all uncertain. Further, the propagation of the waves also depends on the topography of the ocean floor, which is also uncertain.

#### 5.3.1 Model

The problem is governed by the following set of equations:

$$\frac{D v_1}{D t} = f_C v_2 - g \frac{\partial h}{\partial x_1} - b v_1 + S_{v_1}, \quad (48)$$

$$\frac{D v_2}{D t} = -f_C v_1 - g \frac{\partial h}{\partial x_2} - b v_2 + S_{v_2}, \quad (49)$$

$$\frac{\partial h}{\partial t} = -\frac{\partial (v_1 (H + h))}{\partial x_1} - \frac{\partial (v_2 (H + h))}{\partial x_2} + S_h, \quad (50)$$

where  $(v_1(\mathbf{x}, \boldsymbol{\xi}, t), v_2(\mathbf{x}, \boldsymbol{\xi}, t))$  is the velocity vector at the surface,  $\mathbf{x} = (x_1, x_2) \in \Omega \subset \mathbb{R}^2$ ,  $h(\mathbf{x}, \boldsymbol{\xi}, t)$  the elevation of the surface from its position at rest,  $H(\mathbf{x}, \boldsymbol{\xi})$  the sea depth,  $f_C$  models the Coriolis force,  $b$  is the viscous drag coefficient,  $g$  the gravity constant and  $S_{v_1}(\mathbf{x}, \boldsymbol{\xi}, t)$ ,  $S_{v_2}(\mathbf{x}, \boldsymbol{\xi}, t)$ ,  $S_h(\mathbf{x}, \boldsymbol{\xi}, t)$  are the source fields. Without loss of generality, the drag  $b$  and the Coriolis force  $f_C$  are neglected. No slip boundary conditions apply for the velocity. The sources are modeled as acting on  $h$  only,  $S_{v_1} \equiv 0$  and  $S_{v_2} \equiv 0$ .  $S_h$  models the source term acting on  $h$  due to, say, an underwater seismic event. The fluid density and the free surface pressure are implicitly assumed constant.

#### 5.3.2 Sources of uncertainty

Letting  $\boldsymbol{\xi} = (\boldsymbol{\xi}', \boldsymbol{\xi}'')$ , the ocean floor topography  $H(\mathbf{x}, \boldsymbol{\xi})$  is described with a  $N_H$ -term expansion:

$$H(\mathbf{x}, \boldsymbol{\xi}') = \overline{H}(\mathbf{x}) + \sum_{i=1}^{N_H} \sqrt{\lambda_i} \xi'_i(\theta) \varphi_i^H(\mathbf{x}), \quad (51)$$

with  $\boldsymbol{\xi}' = (\xi'_1 \dots \xi'_{N_H})$  the stochastic germ associated to the uncertainty in  $H$ . Random variables  $\{\xi'_i\}_{i=1}^{N_H}$  are *iid*, uniformly distributed. The source  $S_h$  is also uncertain and is modeled as a time-dependent, spatially distributed, quantity:

$$S_h(\mathbf{x}, \boldsymbol{\xi}'', t) = a_t(t) a_{\xi}(\xi''_1) \exp \left( -\frac{(\mathbf{x} - \mathbf{x}_{S_h}(\xi''_3))^T (\mathbf{x} - \mathbf{x}_{S_h}(\xi''_3))}{\sigma_{S_h}(\xi''_2)^2} \right), \quad (52)$$

where  $a_t(t)$  is a given time envelop,  $a_\xi(\xi_1'')$  the uncertain source magnitude,  $\sigma_{S_h}(\xi_2'')$  drives the uncertain source spatial extension and  $\mathbf{x}_{S_h}(\xi_3'')$  is the uncertain spatial location.

The solution of the Shallow Water problem then lies in a  $(d = N_H + 3)$ -dimensional stochastic space. Full details on the problem and the numerical implementation of the simulation are given in Mathelin *et al.* (2011).

### 5.3.3 Approximation from an available database

As an illustration of the methodology, we aim at approximating the sea surface field at a fixed amount of time  $t^*$  after a seismic event. The QoI is then a random field  $u(\mathbf{x}, \boldsymbol{\xi}) = h(\mathbf{x}, \boldsymbol{\xi}, t^*)$ . An accurate description of this field is of importance for emergency plans in case of a seaquake. Sea level measurements of the surface at various spatial locations from past events constitute the experimental dataset  $\{\mathbf{x}^{(q)}, \boldsymbol{\xi}^{(q)}, h(\mathbf{x}^{(q)}, \boldsymbol{\xi}^{(q)}, t^*)\}_{q=1}^{N_q}$  used to derive an approximation of  $u$  under a separated form:  $u(\mathbf{x}, \boldsymbol{\xi}) \approx \langle u \rangle_{N_q}(\mathbf{x}) + \sum_{n=1}^N w_n(\mathbf{x}) \lambda_n(\boldsymbol{\xi})$ .

The solution method here relies on a  $N_q = 37,000$ -sample dataset complemented with  $\widehat{N}_q = 5000$  cross-validation samples and a  $\widetilde{N}_q = 5000$  set for error estimation. We consider a  $N_H = 5$  expansion for the topography, leading to a stochastic dimension of  $d = 5 + 3 = 8$ . The effective number of samples per dimension is then about  $N_q^{1/(d_{\mathbf{x}}+d)} \simeq 2.9$ . The approximation is determined based on a  $|\mathcal{J}_x| = 484$  spatial discretization DOFs (spectral elements) at the deterministic level and  $p = 6$ -th order Legendre polynomials  $\{\psi_\alpha\}$ ,  $N_l = 3$ ,  $N_l^{(\text{PC})} = 3$ , for the stochastic modes. The cardinality of this *a priori* basis is then  $|\mathcal{J}_x| |\mathcal{J}_{\text{prior}}| \simeq 770 \times 10^3 \gg N_q$ , again relying on an efficient subset selection step to make the approximation problem well-posed.

The approximation error when the rank  $N$  varies is shown in Table 3. It is seen that estimating the mean spatial mode  $w_0$  leads to a relative error of about 0.12 while adding the first  $(w_1, \lambda_1)$  and second  $(w_2, \lambda_2)$  pair drops it to about 0.05. Further adding pairs does not lower the approximation error with this dataset and more samples would be needed to accurately estimate them. Spatial modes  $w_0$  and  $w_1$  of the separated approximation are plotted in Fig. 10 for illustration.

$N$	0	1	2	3
$\varepsilon$	0.117	0.056	0.046	0.044

Table 3

Relative approximation error  $\varepsilon$  evolution with the decomposition rank  $N$ .  $N_q = 37,000$ .

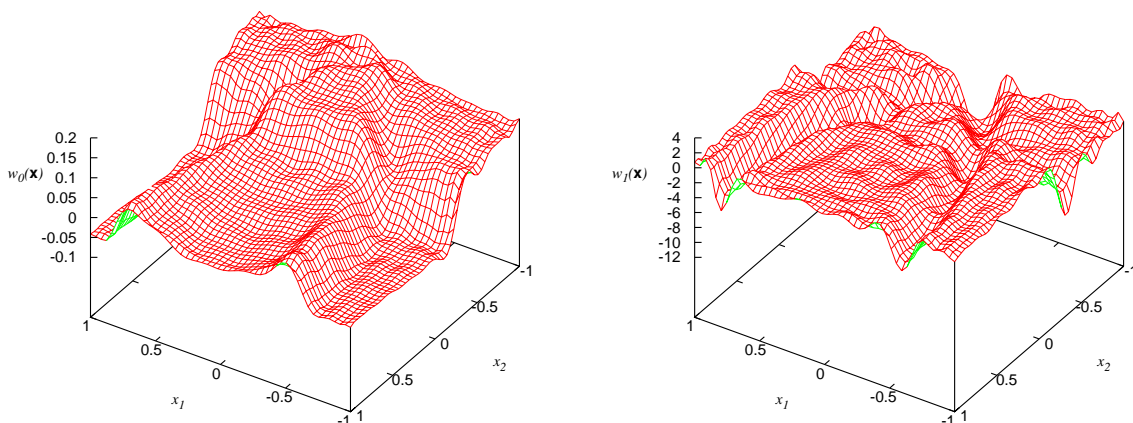


Fig. 10. Mean ( $w_0(\mathbf{x}) \equiv \langle u \rangle_{N_q}(\mathbf{x})$ , left) and first ( $w_1(\mathbf{x})$ , right) spatial modes.

## 6 Conclusion

In this paper, a methodology was put together for deriving a functional representation of a random process only known through a collection of its pointwise evaluations. The proposed method essentially relies on an efficient determination of an approximation basis consistent with the available information. This involves the choice of an *a priori* canonical HDMR format combined with tuning the basis *via* a data-driven subset selection step. This subset selection is carried-out in a bottom-to-top manner, as opposed to a top-to-bottom manner as done in the Compressed Sensing standard framework. It essentially sorts the approximation HDMR modes (groups of predictors) by their contribution in approximating the Quantity of Interest. The final approximation can rely on a different functional description of the modes, typically of higher order and/or nonlinear in the coefficients. This allows to take advantage of both a computationally efficient subset selection process, allowing for a relevant basis structure given the data, and a parsimonious final approximation basis relying on a fewer number of coefficients hence making the most out of the data.

The method is progressive and data-driven. Its efficiency was demonstrated on two examples which have shown its ability to achieve a good approximation accuracy from a small dataset, as long as the quantity at hand is essentially lying on a low-dimensional manifold. In particular, the dominant dimensions are naturally revealed so that all the available information can be dedicated to approximate relevant dependences only. Through a total least squares approach, it was also shown that some robustness can be achieved, an important feature when dealing with experimental data. Using a robust approximation was shown to bring up to a 2-fold improvement upon the approximation error using standard least squares, but at the price of a computational overhead. The global solution method scales reasonably well, exhibiting a linear dependence with the cardinality of the proposal basis dictionary and a quadratic or cubic dependence with the number of samples, depending on the coefficients update strategy.



The present work was focused on a general methodology, disregarding fine-tuning aspects. Among other things, a natural improvement would be to carry-out a predictor selection within each retained groups  $\{f_{\gamma \in \mathcal{J}_{f,\text{post}}}\}$ , further lowering the number of coefficients involved in the approximation and hence allowing, say, higher polynomial orders to be used with the same amount of data. Further, the tensor structure of the Hilbert stochastic space can be exploited and developments towards a data-driven multilinear algebra effective tool for high-dimensional uncertainty quantification are currently carried-out.

## Acknowledgement

The author gratefully acknowledges Tarek El Moselhy and Faidra Stavropoulou for stimulating discussions and useful comments. This work is part of the TYCHE project (ANR-2010-BLAN-0904) supported by the French Research National Agency (ANR).

## References

- ALIŞ O. & RABITZ H., Efficient implementation of high-dimensional model representations, *J. Math. Chem.*, **29** (2), p. 127–142, 2001.
- BERVEILLER M., SUDRET B. & LEMAIRE M., Stochastic finite element: a non intrusive approach by regression, *J. Eur. Meca. Num.*, **15** (1-2-3), p. 81–92, 2006.
- BLATMAN G. & SUDRET B., Adaptive sparse polynomial chaos expansion based on least angle regression, *J. Comput. Phys.*, **230** (6), p. 2345–2367, 2011.
- CAI T.T., WANG L. & XU G., New bounds for restricted isometry constants, *IEEE Trans. Infor. Theo.*, **56** (9), p. 4388–4394, 2010.
- CANDÈS E.J. & ROMBERG J., Sparsity and incoherence in compressive sampling, *Inv. Problems*, **23**, p. 969–985, 2006.
- CANDÈS E.J. & TAO T., Decoding by linear programming, *IEEE Trans. Inform. Theory*, **51**, p. 4203–4215, 2004a.
- CANDÈS E.J. & TAO T., Near-optimal signal recovery from random projections: universal encoding strategies, *IEEE Trans. Inform. Theory*, **52**, p. 5406–5425, 2004b.
- CHEN S., DONOHO D. & SAUNDERS M., Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, **20**, p. 33–61, 1999.
- CHOI S.-K., GRANDHI R.V., CANFIELD R.A. & PETTIT C.L., Polynomial chaos expansion with Latin hypercube sampling for estimating response variability, *AIAA J.*, **42** (6), p. 1191–1198, 2004.
- DEVORE R., PETROVA G. & WOJTASZCZYK P., Approximation of functions of few variables in high dimensions, *Constr. Approx.*, **33**, p. 125–143, 2011.
- DONOHO D.L., Compressed sensing, *IEEE Trans. Infor. Theo.*, **52** (4), p. 1289–1306, 2006.

- DOOSTAN A. & IACCARINO G., A least-squares approximation of partial differential equations with high-dimensional random inputs, *J. Comput. Phys.*, **228**, p. 4332–4345, 2009.
- DOOSTAN A. & OWHADI H., A non-adapted sparse approximation of PDEs with stochastic inputs, *J. Comput. Phys.*, **230** (8), p. 3015–3034, 2011.
- EFRON B., HASTIE T., JOHNSTONE I. & TIBSHIRANI R., Least Angle Regression, *Annals of Statistics*, **32**, p. 407–499, 2004.
- GANAPATHYSUBRAMANIAN B. & ZABARAS N., Sparse grid collocation methods for stochastic natural convection problems, *J. Comput. Phys.*, **225**, p. 652–685, 2007.
- GHANEM R.G. & SPANOS P.D., *Stochastic finite elements. A spectral approach*, rev. edn., Springer Verlag, 222 p., 2003.
- HASTIE T., TIBSHIRANI R. & FRIEDMAN J., *The Elements of Statistical Learning*, 2nd edn., Springer Verlag, 745 p., 2009.
- HESTERBERG T., CHOI N. H., MEIER L. & FRALEY C., Least angle and  $\ell_1$  penalized regression: A review, *Statistics Surveys*, **2**, p. 61–93, 2008.
- HOMMA T. & SALTELLI A., Importance measures in global sensitivity analysis of non-linear models, *Rel. Engrg. Sys. Safety*, **52** (1), p. 1–17, 1996.
- KHOROMSKIJ B.N., Tensors-structured numerical methods in scientific computing: Survey on recent advances, *Chemometrics and Intelligent Laboratory Systems*, **110** (1), p. 1–19, 2012.
- KHOROMSKIJ B.N. & SCHWAB C., Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs, *SIAM J. Sci. Comput.*, **33** (1), p. 1–25, 2011.
- KUO F.Y., SLOAN I.H., WASILKOWSKI G.W. & WOŹNIAKOSKI H., On decompositions of multivariate functions, *Mathematics of Computation*, **79** (270), p. 953–966, 2009.
- LE MAÎTRE O.P. & KNIO O.M., *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, Springer, 552 p., 2010.
- MA X. & ZABARAS N., An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations, *J. Comput. Phys.*, **10**, p. 3884–3915, 2010.
- MARKOVSKY I. & VAN HUFFEL S., Overview of total least squares methods, *Signal Processing*, **87** (10), p. 2283–2302, 2007.
- MATHELIN L., DESCELIERS C. & HUSSAINI M.Y., Stochastic data assimilation with a Polynomial Chaos parametric estimation, *Comp. Mech.*, **47** (6), p. 603–616, 2011.
- MATHELIN L. & GALLIVAN K.A., A compressed sensing approach for partial differential equations with random input data, *Comm. Comp. Phys.*, **12**, p. 919–954, 2012.
- MATTHIES H.G. & ZANDER E., Solving stochastic systems with low-rank tensor compression, *Linear Algebra and its Applications*, **436** (10), p. 3819–3838, 2012.
- NOBILE F., TEMPONE R. & WEBSTER C., An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, *SIAM J. Numer. Anal.*, **46** (5), p. 2411–2442, 2007.
- NOUY A., A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations, *Comput. Methods Appl. Mech. Engrg.*, **196** (45–48), p. 4521–4537, 2007.
- NOUY A., A priori model reduction through Proper Generalized Decomposition for solving time-dependent partial differential equations, *Comput. Methods Appl. Mech. Engrg.*,

- 199**, p. 1603–1626, 2010*a*.
- NOUY A., Proper Generalized Decompositions and separated representations for the numerical solution of high dimensional stochastic problems, *Archives of Computational Methods in Engineering*, **17**, p. 403–434, 2010*b*.
- NOVAK E. & RITTER K., Simple cubature formulas with high polynomial exactness, *Constructive Approximation*, **15**, p. 499–522, 1999.
- RABITZ H. & ALIŞ O., General foundations of high-dimensional model representations, *J. Math. Chem.*, **25**, p. 197–233, 1999.
- SOBOL I.M., Sensitivity estimates for nonlinear mathematical models, *Mathematical Modelling and Computational Experiments*, **1** (4), p. 407–414, 1993.
- SOIZE C. & GHANEM R., Physical systems with random uncertainties: chaos representations with arbitrary probability measure, *SIAM J. Sci. Comput.*, **26** (2), p. 395–410, 2004.
- TIBSHIRANI R., Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B*, **58**, p. 267–288, 1996.
- WENTZELL P.D., ANDREWS D.T., HAMILTON D.C., FABER K. & KOWALSKI B.R., Maximum likelihood principal component analysis, *J. Chemometrics*, **11**, p. 339–366, 1997.
- WIENER N., The homogeneous chaos, *Amer. J. Math.*, **60** (4), p. 897–936, 1938.
- XIE J. & ZENG L., 2010 Group Variable Selection Methods and Their Applications in Analysis of Genomic Data. In *Frontiers in Computational and Systems Biology, Computational Biology* (ed. J. Feng, W. Fu & F. Sun), , vol. 15, p. 231–248. Springer-Verlag.
- XIU D. & HESTHAVEN J., High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, **27**, p. 1118–1139, 2005.
- XIU D. & KARNIADAKIS G.E., The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM J. Sci. Comput.*, **24** (2), p. 619–644, 2002.
- YUAN M. & LIN Y., Model selection and estimation in regression with grouped variables, *J. R. Statist. Soc. B*, **68**, p. 49–67, 2006.